GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: June 14, 2003, 06:36:14 ; Search time 2944 Seconds
(without alignments)
10369.852 Million cell updates/sec

Title: US-09-896-186B-23
Perfect score: 1049
Sequence: 1 accaaagcattaatttttat.....tttaaccgtccagaaactag 1049

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1049	100.0	1049	6	AX375195	AX375195 Sequence
2		1015	96.8	1137	8	ATH404476	AJ404476 Arabidops
3		831	79.2	942	6	AX375173	AX375173 Sequence
4	c	302.2	28.8	118718	8	ATF18A5	AL035528 Arabidops
5	c	302.2	28.8	199667	8	ATCHRIV37	AL161537 Arabidops
6	c	151.2	14.4	130299	8	AC019012	AC019012 Genomic S
7		142	13.5	82484	8	AC002341	AC002341 Arabidops
8		111	10.6	4299	6	AX375189	AX375189 Sequence
9		111	10.6	5189	6	AX333649	AX333649 Sequence
10		111	10.6	5189	6	E16596	E16596 Human WS mR
11		111	10.6	5189	9	HUMDR	L76937 Homo sapien
12		111	10.6	5208	9	AF091214	AF091214 Homo sapi
13		110.6	10.5	5143	5	AF067418	AF067418 Xenopus l
14	c	97.8	9.3	200576	8	ATFCA0	Z97335 Arabidopsis
15		82.8	7.9	6476	6	AR149285	AR149285 Sequence
16		82.8	7.9	6476	10	AF241636	AF241636 Mus muscu
17		81.8	7.8	4201	10	MMU97045	U97045 Mus musculu
18		81.2	7.7	4955	10	D86527	D86527 Mouse werne
19		81.2	7.7	5034	10	D86526	D86526 Mouse werne
20		81.2	7.7	5050	10	AF091215	AF091215 Mus muscu
21		81.2	7.7	5058	6	E16141	E16141 cDNA encodi
22	c	65.4	6.2	155597	8	OSJN00161	AL662964 Oryza sat
23		55.4	5.3	473	11	G73264	G73264 csnpwrn-pcr
24		55.4	5.3	98844	9	HSWRNG1	AF181896 Homo sapi
25		55.4	5.3	180098	9	AC084736	AC084736 Homo sapi
26		51.2	4.9	97348	10	AF091216	AF091216 Mus muscu
27	c	51.2	4.9	227725	2	AC115809	AC115809 Mus muscu
28		49.2	4.7	7218	6	I66494	I66494 Sequence 14
29	c	48.2	4.6	174993	2	AC098510	AC098510 Rattus no
30		47	4.5	168698	2	OSJN00023	AL606588 Oryza sat
31		46	4.4	2865	10	BC018508	BC018508 Mus muscu
32		45.2	4.3	434	11	G73268	G73268 csnpwrn-pcr
33		45	4.3	236120	14	AF063866	AF063866 Melanoplu
34		44.4	4.2	2077	9	BC001962	BC001962 Homo sapi
35		44.4	4.2	2833	9	AK001600	AK001600 Homo sapi
36		44.4	4.2	2844	9	HSM805527	AL834434 Homo sapi
37		44.4	4.2	3404	9	AK056025	AK056025 Homo sapi
38		44.4	4.2	198421	2	AC125180	AC125180 Mus muscu
39	c	44.2	4.2	164679	2	AC096680	AC096680 Pan trogl
40		43.4	4.1	125020	9	AF429315	AF429315 Homo sapi
41	c	43	4.1	115489	2	AC117072	AC117072 Dictyoste
42	c	42.8	4.1	408	6	AX300942	AX300942 Sequence
43	c	42.6	4.1	116660	9	AC002074	AC002074 Homo sapi
44		42.4	4.0	160780	2	AL445653	AL445653 Homo sapi
45		42.2	4.0	1141	6	AX083744	AX083744 Sequence

ALIGNMENTS

RESULT 1
AX375195
LOCUS AX375195 1049 bp DNA linear PAT 01-MAR-2002
DEFINITION Sequence 23 from Patent WO0210362.
ACCESSION AX375195
VERSION AX375195.1 GI:19169935
KEYWORDS .
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Levin,J.Z., Phillips,K.L., Budziszewski,G.J., Meins,F.J. and
Glazov,E.A.

```
  TITLE     Methods of controlling gene expression and gene silencing
  JOURNAL   Patent: WO 0210362-A 23 07-FEB-2002;
            Syngenta Participations AG (CH) ; Novartis Forschungsstiftung
            Zweigniederlassung (CH)
FEATURES             Location/Qualifiers
     source          1. .1049
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
BASE COUNT      292 a    216 c    220 g    321 t
ORIGIN

  Query Match           100.0%;  Score 1049;  DB 6;  Length 1049;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-259;
  Matches 1049;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGA 60

Qy         61 TCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACA 120

Qy        121 ATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCT 180

Qy        181 CCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGC 240

Qy        241 CTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA 300

Qy        301 GGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGG 360

Qy        361 TTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAA 420

Qy        421 TAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGG 480

Qy        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540

Qy        541 TTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGG 600

Qy        601 TAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCA 660

Qy        661 AAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGG 720

Qy        721 GCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCA 780

Qy        781 GGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGG 840

Qy        841 ATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTG 900

Qy        901 GCTCATAACGTGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCTCATAACGTGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAA 960

Qy        961 ATGATATGATACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATATGATACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGT 1020

Qy       1021 TCTGGTCCCTTTAACCGTCCAGAAACTAG 1049
              |||||||||||||||||||||||||||||
Db       1021 TCTGGTCCCTTTAACCGTCCAGAAACTAG 1049


RESULT 2
ATH404476
LOCUS       ATH404476               1137 bp    mRNA    linear   PLN 02-NOV-2000
DEFINITION  Arabidopsis thaliana mRNA for exonuclease (wrnexo gene).
ACCESSION   AJ404476
VERSION     AJ404476.1  GI:11121454
KEYWORDS    exonuclease; wrnexo gene.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1137)
  AUTHORS   Hartung,F., Plchova,H. and Puchta,H.
  TITLE     Molecular characterisation of RecQ homologues in Arabidopsis
            thaliana
  JOURNAL   Nucleic Acids Res. 21, 4275-4282 (2000)
REFERENCE   2  (bases 1 to 1137)
  AUTHORS   Hartung,F., Plchova,H. and Puchta,H.
  TITLE     Molecular characterisation of RecQ homologues in Arabidopsis
            thaliana
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1137)
  AUTHORS   Hartung,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2000) Hartung F., Cytogenetics, Institute of
            plant genetics and crop plant research (IPK), Corrensstr. 3,
            Sachsen Anhalt, 06466 Gatersleben, GERMANY
  REMARK    Revised by author 14-JUN-2000
COMMENT     Related to AL161537 and AL035528.
FEATURES             Location/Qualifiers
     source          1. .1137
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /tissue_type="flowers"
     gene            36. .893
                     /gene="wrnexo"
     CDS             36. .893
                     /gene="wrnexo"
                     /function="putative role in replication and recombination"
                     /codon_start=1
                     /product="exonuclease"
                     /protein_id="CAC14871.1"
                     /db_xref="GI:11121455"
                     /translation="MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQA
                     TTSVHGHEEDPNQIPNNIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSK
                     TATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNY
                     CDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQ
                     KIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYK
                     DLPDAVSGS"
BASE COUNT      318 a    226 c    238 g    355 t
ORIGIN

  Query Match           96.8%;  Score 1015;  DB 8;  Length 1137;
  Best Local Similarity   99.1%;  Pred. No. 7.6e-251;
  Matches 1034;  Conservative    0;  Mismatches    0;  Indels    9;  Gaps    1;
```

```
Qy          7 GCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATCGACG 66
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATCGACG 60

Qy         67 ACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCT 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCT 120

Qy        127 CCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCC 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCC 180

Qy        187 ATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTT 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTT 240

Qy        247 CCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATT 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATT 300

Qy        307 TTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATA 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATA 360

Qy        367 AGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTT 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTT 420

Qy        427 TTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGG 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGG 480

Qy        487 TTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATT 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATT 540

Qy        547 CTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTA 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTA 600

Qy        607 TTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATG 660

Qy        667 TTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTG 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTG 720

Qy        727 CCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTG 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTG 780

Qy        787 GGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTT 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTT 840

Qy        847 ATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTGGCTCAT 906
              ||||||||||||||||||||          ||||||||||||||||||||||||||||||
Db        841 ATGCTTCATGGCATCTTTAC---------AAGGACCTTCCTGATGCTGTCAGTGGCTCAT 891

Qy        907 AACGTGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAAATGATA 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        892 AACGTGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAAATGATA 951

Qy        967 TGATACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGTTCTGGT 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        952 TGATACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGTTCTGGT 1011

Qy       1027 CCCTTTAACCGTCCAGAAACTAG 1049
              |||||||||||||||||||||||
Db       1012 CCCTTTAACCGTCCAGAAACTAG 1034
```

```
RESULT 3
AX375173
LOCUS       AX375173                942 bp    DNA     linear   PAT 01-MAR-2002
DEFINITION  Sequence 1 from Patent WO0210362.
ACCESSION   AX375173
VERSION     AX375173.1  GI:19169923
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Levin,J.Z., Phillips,K.L., Budziszewski,G.J., Meins,F.J. and
            Glazov,E.A.
  TITLE     Methods of controlling gene expression and gene silencing
  JOURNAL   Patent: WO 0210362-A 1 07-FEB-2002;
            Syngenta Participations AG (CH) ; Novartis Forschungsstiftung
            Zweigniederlassung (CH)
FEATURES             Location/Qualifiers
     source          1. .942
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
BASE COUNT      263 a    203 c    195 g    281 t
ORIGIN

  Query Match           79.2%;  Score 831;  DB 6;  Length 942;
  Best Local Similarity   96.1%;  Pred. No. 2.1e-203;
  Matches  852;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         42 ATGTCATCGTCAAATTGGATCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGAC 101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCATCGTCAAATTGGATCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGAC 60

Qy        102 GCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCG 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCG 120

Qy        162 ACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAAT 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAAT 180

Qy        222 AATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCT 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCT 240

Qy        282 CTCTCCCGTTGCCGAGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTAT 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCTCCCGTTGCCGAGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTAT 300

Qy        342 AGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACC 401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACC 360

Qy        402 AAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTT 461
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTT 420

Qy        462 AGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTAT 521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTAT 480

Qy        522 TGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAA 581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAA 540

Qy        582 GATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCAT 641
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCAT 600

Qy        642 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 701
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        601 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 660

Qy        702 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 720

Qy        762 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 780

Qy        822 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGAC 881
              ||||||||||||||||||||||||||||||||||||||||||||||    |
Db        781 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTAACAACAACG 840

Qy        882 CTTCCTGATGCTGTCAGTGGCTCATAACGTGAAGGAGGAAGCTTAAA 928
                  ||  | |||    |    | | ||||| |  || || | |
Db        841 AAAAACCATCTTCTCACACTCAACGACCTTGAAGCAAAAATCTCACA 887


RESULT 4
ATF18A5/c
LOCUS       ATF18A5               118718 bp    DNA     linear   PLN 18-AUG-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone F18A5 (ESSA
            project).
ACCESSION   AL035528
VERSION     AL035528.2  GI:5748493
KEYWORDS    .
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 118718)
  AUTHORS   Bevan,M., Weber,N., Grueninger,D., Schmidheini,T., Bancroft,I.,
            Mewes,H.W., Mayer,K.F.X., Lemcke,K. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 118718)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     On Aug 20, 1999 this sequence version replaced gi:4455290.
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .118718
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     misc_feature    1. .3993
                     /note="basepairs 1-3993 are not integral part of clone
                     F18A5 but are part ofBAC FT23H4"
     gene            3154. .4320
                     /gene="F18A5.10"
     CDS             3154. .4320
                     /gene="F18A5.10"
                     /note="weak similarity to CADMIUM-INDUCED PROTEIN AS30 -
                     Arabidopsis thaliana, GB:P42736"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB36826.2"
                     /db_xref="GI:5748494"
                     /translation="MITPIHTQHSLILVYINIYSPPILSKLRTGFILWTNTQKTNKKR
                     NMEDQFPKIETSFMHDKLLSSGIYGFLSSSTPPQLLGVPIFLEGMKSPLLPASSTPSY
                     FVSPHDHELTSSIHPSPVASVPWNFLESFPQSQHPDHHPSKPPNLTLFLKEPKLLELS
                     QSESNMSPYHKYIPNSFYQSDQNRNEWVEINKTLTNYPSKGFGNYWLSTTKTQPMKSK
                     TRKVVQTTTPTKLYRGVRQRHWGKWVAEIRLPRNRTRVWLGTFETAEQAAMAYDTAAY
                     ILRGEFAHLNFPDLKHQLKSGSLRCMIASLLESKIQQISSSQVSNSPSPPPPKVGTPE
                     QKNHHMKMESGEDVMMKKQKSHKEVMEGDGVQLSRMPSLDMDLIWDALSFPHSS"
     exon            3154. .4320
                     /gene="F18A5.10"
                     /number=1
     gene            5109. .7045
                     /gene="F18A5.20"
     gene            complement(join(5109. .5150,5302. .6455,6496. .7045))
                     /gene="F18A5.20"
     CDS             complement(join(5109. .5150,5302. .6455,6496. .7045))
                     /gene="F18A5.20"
                     /note="similarity to IFA-binding protein, Arabidopsis
                     thaliana, AF004556"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB36827.1"
                     /db_xref="GI:4455292"
                     /translation="MRCQEVKSWTFKGLVAAFIDLSVAFSLLCASFIVYVTSKFLGLF
                     GLDLPCPCDGLYSECFQESLRNLPVKKISSVQRSVKNRTPFDSILYNGGKKRKCERRR
                     VQLEDEVSSTTPSVGKFENKASGFDLLTAQSLKKGSFKVKSKRLSFHRSPYGFKNHYQ
                     SCLGLKSFEGSYDENDPLLVNSNVNVDFDASCNIVDSGKALEDVSLRKSVSLSSVGCG
                     VGEGACSSPGMVQRTVEMSEQVLGEERAARASLALELEKERNAAASAADEALGMILRL
                     QEEKASIEMEARQYQRMIEEKSAFDAEEMSILKEILLRREREKHFLEKEVDTYRQMFL
                     ETEQPHNTPDSKPAQIERLQTPQQITEPWDDMETVNVSSGFEIFTNQMDNTECDQSRS
                     EPFDALSDIELKEREEGETLYAELVSRTSDIAVSKKLCEDPHDIDCHVHDIHVVTDED
                     NKGQLNVPSDHATQDLKLDRSQSVSDTSYAFPQGKSNMSTNMRRNSMSAIDYERLKIE
                     SEVGLLRGRLRAVQKGREKISFSSKDQSKSQVQRCGDKTSRFWEARRSGHLDSSSPSS
                     TMVKAMSMPLDLHSA"
     exon            complement(5109. .5150)
                     /gene="F18A5.20"
                     /number=1
     intron          complement(5151. .5301)
                     /gene="F18A5.20"
                     /number=1
     exon            complement(5302. .6455)
                     /gene="F18A5.20"
                     /number=2
     intron          complement(6456. .6495)
                     /gene="F18A5.20"
                     /number=2
     exon            complement(6496. .7045)
                     /gene="F18A5.20"
                     /number=3
     gene            7880. .9513
                     /gene="F18A5.30"
     gene            complement(join(7880. .8293,8499. .8568,8647. .8725,
                     8918. .9055,9232. .9255,9330. .9513))
                     /gene="F18A5.30"
     CDS             complement(join(7880. .8293,8499. .8568,8647. .8725,
                     8918. .9055,9232. .9255,9330. .9513))
                     /gene="F18A5.30"
                     /note="contains EST gb:H76035, AA651479"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB36828.1"
                     /db_xref="GI:4455293"
                     /translation="MYSAIRSSLPLDGSLGDYSDGTNLPIDACLVLTTDPKPRLRWTS
                     ELHERFVDAVTQLGGPDISTKNLTTKATPKTIMRTMGVKGLTLYHLKSHLQVLMLLPS
                     ISNYASLATYNKQISCVAESQDTGSSSTSSLRLAAQEQNESYQVTEALRAQMEVQRRL
                     HEQLEVQRRLQLRIEAQGKYLQSILEKACKAIEEQAVAFAGLEAAREELSELAIKASI
                     TNGCQGTTSTFDTTKMMIPSLSELAVAIEHKNNCSAESSLTSSTVGSPVSAALMKKRQ
                     RGVFGNGDSVVVGHDAGWVMPSSSIG"
     exon            complement(7880. .8293)
                     /gene="F18A5.30"
                     /number=1
     intron          complement(8294. .8498)
                     /gene="F18A5.30"
                     /number=1
     exon            complement(8499. .8568)
                     /gene="F18A5.30"
                     /number=2
     intron          complement(8569. .8646)
                     /gene="F18A5.30"
```

```
                /number=2
exon            complement(8647. .8725)
                /gene="F18A5.30"
                /number=3
intron          complement(8726. .8917)
                /gene="F18A5.30"
                /number=3
exon            complement(8918. .9055)
                /gene="F18A5.30"
                /number=4
intron          complement(9056. .9231)
                /gene="F18A5.30"
                /number=4
exon            complement(9232. .9255)
                /gene="F18A5.30"
                /number=5
intron          complement(9256. .9329)
                /gene="F18A5.30"
                /number=5
exon            complement(9330. .9513)
                /gene="F18A5.30"
                /number=6
gene            10627. .13823
                /gene="F18A5.40"
gene            complement(join(10627. .13689,13812. .13823))
                /gene="F18A5.40"
CDS             complement(join(10627. .13689,13812. .13823))
                /gene="F18A5.40"
                /note="similarity to various predicted proteins,
                Arabidopsis thaliana
                Contains Cytochrome c family heme-binding site signature
                [CNDCHA]"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB36829.1"
                /db_xref="GI:4455294"
                /translation="MVFLTLCGTRRASFAAISVYISEDESFQEKRIDSVENRGIRPNH
                QTLKWLLEGCLKTNGSLDEGRKLHSQILKLGLDSNGCLSEKLFDFYLFKGDLYGAFKV
                FDEMPERTIFTWNKMIKELASRNLIGEVFGLFVRMVSENVTPNEGTFSGVLEACRGGS
                VAFDVVEQIHARILYQGLRDSTVVCNPLIDLYSRNGFVDLARRVFDGLRLKDHSSWVA
                MISGLSKNECEAEAIRLFCDMYVLGIMPTPYAFSSVLSACKKIESLEIGEQLHGLVLK
                LGFSSDTYVCNALVSLYFHLGNLISAEHIFSNMSQRDAVTYNTLINGLSQCGYGEKAM
                ELFKRMHLDGLEPDSNTLASLVVACSADGTLFRGQQLHAYTTKLGFASNNKIEGALLN
                LYAKCADIETALDYFLETEVENVVLWNVMLVAYGLLDDLRNSFRIFRQMQIEEIVPNQ
                YTYPSILKTCIRLGDLELGEQIHSQIIKTNFQLNAYVCSVLIDMYAKLGKLDTAWDIL
                IRFAGKDVVSWTTMIAGYTQYNFDDKALTTFRQMLDRGIRSDEVGLTNAVSACAGLQA
                LKEGQQIHAQACVSGFSSDLPFQNALVTLYSRCGKIEESYLAFEQTEAGDNIAWNALV
                SGFQQSGNNEEALRVFVRMNREGIDNNNFTFGSAVKAASETANMKQGKQVHAVITKTG
                YDSETEVCNALISMYAKCGSISDAEKQFLEVSTKNEVSWNAIINAYSKHGFGSEALDS
                FDQMIHSNVRPNHVTLVGVLSACSHIGLVDKGIAYFESMNSEYGLSPKPEHYVCVVDM
                LTRAGLLSRAKEFIQEMPIKPDALVWRTLLSACVVHKNMEIGEFAAHHLLELEPEDSA
                TYVLLSNLYAVSKKWDARDLTRQKMKEKGVKKEPGQSWIEVKNSIHSFYVGDQNHPLA
                DEIHEYFQDLTKRASEIGYVQDCFSLLNELQHEQKDPIIFIHSEKLAISFGLLSLPAT
                VPINVMKNLRVCNDCHAWIKFVSKVSNREIIVRDAYRFHHFEGGACSCKDYW"
exon            complement(10627. .13689)
                /gene="F18A5.40"
                /number=1
intron          complement(13690. .13811)
                /gene="F18A5.40"
                /number=1
exon            complement(13812. .13823)
                /gene="F18A5.40"
                /number=2
gene            17299. .19268
                /gene="F18A5.50"
CDS             join(17299. .17646,18255. .18495,18615. .18818,19108. .19268)
                /gene="F18A5.50"
                /note="strong similarity to isoflavone reductase-like
                protein, Lupinus albus, gb:U48590
                contains EST gb:T21063, T88563, H36393, AA720215, T76609,
                H36674"
                /codon_start=1
                /product="isoflavone reductase-like protein"
                /protein_id="CAB36830.1"
                /db_xref="GI:4455295"
```

```
  Query Match             28.8%;  Score 302.2;  DB 8;  Length 118718;
  Best Local Similarity   97.5%;  Pred. No. 3.6e-67;
  Matches  307;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 CCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGAT 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96598 CCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGAT 96539

Qy         62 CGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAA 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96538 CGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAA 96479

Qy        122 TTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTC 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96478 TTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTC 96419

Qy        182 CGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCC 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96418 CGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCC 96359

Qy        242 TCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAG 301
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      96358 TCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGGTAT 96299

Qy        302 GAATTTTCCAGCAAT 316
              ||| ||| ||| |||
Db      96298 GAAACTTCTTCAAAT 96284
```

```
RESULT 5
ATCHRIV37/c
LOCUS       ATCHRIV37              199667 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 37.
ACCESSION   AL161537
VERSION     AL161537.2  GI:7268064
KEYWORDS    .
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 21093 to 139810)
  AUTHORS   Weber,N., Grueninger,D., Schmidheini,T., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199667)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV36 at the 5' end and an
            overlap with ATCHRIV38 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .199667
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            7130. .13339
                     /gene="AT4g13610"
     CDS             join(7130. .7298,7473. .8347,8449. .8516,9309. .10812,
                     10929. .11093,11197. .11380,11465. .11576,11636. .11816,
                     11901. .12098,12231. .12509,12633. .12761,12846. .13061,
```

```
                     13205. .13339)
                     /gene="AT4g13610"
                     /note="strong similarity to DNA
                     (cytosine-5-)-methyltransferase, Arabidopsis thaliana ,
                     PIR2:S59604
                     Contains C-5 cytosine-specific DNA methylases signatures
                     AA1077-1089;C-5 cytosine-specific DNA methylases
                     signatures AA1377-1395
                     contains EST gb:N38410, AI994691.1, R90103, AA605459"
                     /codon_start=1
                     /product="DNA (cytosine-5-)-methyltransferase-like
                     protein"
                     /protein_id="CAB78403.1"
                     /db_xref="GI:7268065"
                     /translation="MKTKAGKQKKRSVDSDDDVSRERRPKRATSGTNFKEKSLRFSEK
                     YETVEAKKEQIVGDDEKEEKGVRFQSFGRVENWTISGYEDGSPVIWISTVIADYDCRK
                     PSKKYKKLYDYFFEKACACVEVCKNLSTNPDTSLKELLAAVVRSMNGRKIFSSGGVIQ
                     EFVISQGEFIYNQLAGLDETSKNHETKFVDNRVLVSLRDESRKIHKAFSNVALRIDES
                     KVLTSDQLMDGGEDEDLKYAKLLQEEEHMKSMDRSRNKRSSTTSAPNKFYIKINEDEI
                     AHDYPLPSYYKNTKDETDELVLFNAGYAVDARNLPCRTLHNWALYNSDLMLISLEFLP
                     MKPCADIDVTYLGQIKEWKIDFGEDMIFVLLRTDMAWYRLGKPSEQYAPWFEPILKTV
                     RIGTSILALLKNETRMAKLSYTDVIKRLCGLEENDQAYISSTFFDVERYVIVHGQIIL
                     QFLTECPDEYIKRCPFVTGLASKMQDRHHTKWIIKKKRKMLQKGENLNLRRGKAPKVS
                     KMKAMQATTTRLINRIWGEFYSIYSPEDPLEEIGAEEEFEEVEDVEEEDENEEEDTIQ
                     KAIEVQKADTLKKIRGSCKEMEIRWEGEILGETCAGEPLYGQALVGGRKMDVGGAVIL
                     EVDDQGETPLIYFVEYMFESSDNSKKLHGKLLQRGSETVLGTAANERELFLTNECLTV
                     QLKDIKGTVSFEIRSRPWGHQYKKEHMAADKLDRARAEERKAKDLPIEYYCKSLYSPE
                     KGGFFSLPRSDMGLGSGFCSSCKIRENEEERSKTKLNDSKTRFLSNGIKYSVGDFVYQ
                     IPNYLSKDRGKRRPVFKYGRNVGLRAFVVCQILDIVDLKEPKKGNTTSFEVKVRRFYR
                     PDDVSAEEAYASDIQEVYYSEDTYILPPEAIKGKCEVMKKTDMPLCREYPILDHVYFC
                     DRFYDSSNGCLKKLPYNMMLKFSTIKDDTLLREKKTETGSAMLLKPDEVPKGKRLATL
                     DIFAGCGGLSYGLEKAGVSDTKWAIEYEEPAAQAFKQNHPKTTVFVDNCNVILRISWL
                     RLLINDRAIMEKCGDVDDCISTTEAAELATKLDENQKSTLPLPGQVDFISGGPPCQGF
                     SRLNRFSDGSWSKNQCQMILAFLSFADYFRPKYFLLENVKTFVSFNEGHTFHLTVASL
                     LEMGYQVRFGLLEAGAYGISQPRKRAFIWAAAPNEVLPEWPEPMHVFNNPGFKIPLSQ
                     GLHYAAVQSTKFGAPFRSITVRDAIGDLPPIESGESKINKEEMRGSMTVLTDHICKKM
                     NELNLIRCKKIPKTPGADWRDLPDEHVNLSNGIVKNIVPNLLNKAKDHNGYKGLYGRL
                     DWHGNLPTCITNLQPMGLVGMCFHPDQDRIISVRECARSQGFPDSYKFSGNIKDKHRQ
                     VGNAVPPPLAFALGRKLKEALHLRNI"
exon                 7130. .7298
                     /gene="AT4g13610"
                     /number=1
intron               7299. .7472
                     /gene="AT4g13610"
                     /number=1
exon                 7473. .8347
                     /gene="AT4g13610"
                     /number=2
intron               8348. .8448
                     /gene="AT4g13610"
                     /number=2
exon                 8449. .8516
                     /gene="AT4g13610"
                     /number=3
intron               8517. .9308
                     /gene="AT4g13610"
                     /number=3
exon                 9309. .10812
                     /gene="AT4g13610"
                     /number=4
intron               10813. .10928
                     /gene="AT4g13610"
                     /number=4
exon                 10929. .11093
                     /gene="AT4g13610"
                     /number=5
intron               11094. .11196
                     /gene="AT4g13610"
                     /number=5
exon                 11197. .11380
                     /gene="AT4g13610"
                     /number=6
intron               11381. .11464
                     /gene="AT4g13610"
                     /number=6
exon                 11465. .11576
                     /gene="AT4g13610"
                     /number=7
intron               11577. .11635
                     /gene="AT4g13610"
                     /number=7
exon                 11636. .11816
                     /gene="AT4g13610"
                     /number=8
intron               11817. .11900
                     /gene="AT4g13610"
                     /number=8
exon                 11901. .12098
                     /gene="AT4g13610"
                     /number=9
intron               12099. .12230
                     /gene="AT4g13610"
                     /number=9
exon                 12231. .12509
                     /gene="AT4g13610"
                     /number=10
intron               12510. .12632
                     /gene="AT4g13610"
                     /number=10
exon                 12633. .12761
                     /gene="AT4g13610"
                     /number=11
intron               12762. .12845
                     /gene="AT4g13610"
                     /number=11
exon                 12846. .13061
                     /gene="AT4g13610"
                     /number=12
intron               13062. .13204
                     /gene="AT4g13610"
                     /number=12
exon                 13205. .13339
                     /gene="AT4g13610"
                     /number=13
misc_feature         21093. .25085
                     /note="basepairs 1-3993 are not integral part of clone
                     F18A5 but are part ofBAC FT23H4"
gene                 24246. .25412
                     /gene="AT4g13620"
CDS                  24246. .25412
                     /gene="AT4g13620"
                     /note="similarity to CADMIUM-INDUCED PROTEIN AS30 -
                     Arabidopsis thaliana, GB:P42736
                     contains EST gb:AW032262.1, AW035882.1, AW004543.1,
                     T45770, AI994647.1, T42962, H37693, N65389, AI994192.1,
                     AA720310, AW029674.1, T76090"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB78404.1"
                     /db_xref="GI:7268066"
                     /translation="MITPIHTQHSLILVYINIYSPPILSKLRTGFILWTNTQKTNKKR
                     NMEDQFPKIETSFMHDKLLSSGIYGFLSSSTPPQLLGVPIFLEGMKSPLLPASSTPSY
                     FVSPHDHELTSSIHPSPVASVPWNFLESFPQSQHPDHHPSKPPNLTLFLKEPKLLELS
                     QSESNMSPYHKYIPNSFYQSDQNRNEWVEINKTLTNYPSKGFGNYWLSTTKTQPMKSK
                     TRKVVQTTTPTKLYRGVRQRHWGKWVAEIRLPRNRTRVWLGTFETAEQAAMAYDTAAY
                     ILRGEFAHLNFPDLKHQLKSGSLRCMIASLLESKIQQISSSQVSNSPSPPPPKVGTPE
                     QKNHHMKMESGEDVMMKKQKSHKEVMEGDGVQLSRMPSLDMDLIWDALSFPHSS"
exon                 24246. .25412
                     /gene="AT4g13620"
                     /number=1
gene                 26201. .28137
                     /gene="AT4g13630"
gene                 complement(join(26201. .26242,26394. .27547,27588. .28137))
                     /gene="AT4g13630"
CDS                  complement(join(26201. .26242,26394. .27547,27588. .28137))
                     /gene="AT4g13630"
```

```
                     /note="similarity to IFA-binding protein, Arabidopsis
                     thaliana, AF004556
                     contains EST gb:N65544"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB78405.1"
                     /db_xref="GI:7268067"
                     /translation="MRCQEVKSWTFKGLVAAFIDLSVAFSLLCASFIVYVTSKFLGLF
                     GLDLPCPCDGLYSECFQESLRNLPVKKISSVQRSVKNRTPFDSILYNGGKKRKCERRR
                     VQLEDEVSSTTPSVGKFENKASGFDLLTAQSLKKGSFKVKSKRLSFHRSPYGFKNHYQ
                     SCLGLKSFEGSYDENDPLLVNSNVNVDFDASCNIVDSGKALEDVSLRKSVSLSSVGCG
                     VGEGACSSPGMVQRTVEMSEQVLGEERAARASLALELEKERNAAASAADEALGMILRL
                     QEEKASIEMEARQYQRMIEEKSAFDAEEMSILKEILLRREREKHFLEKEVDTYRQMFL
                     ETEQPHNTPDSKPAQIERLQTPQQITEPWDDMETVNVSSGFEIFTNQMDNTECDQSRS
                     EPFDALSDIELKEREEGETLYAELVSRTSDIAVSKKLCEDPHDIDCHVHDIHVVTDED

  Query Match             28.8%;  Score 302.2;  DB 8;  Length 199667;
  Best Local Similarity   97.5%;  Pred. No. 3.5e-67;
  Matches  307;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 CCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGAT 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     117690 CCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGAT 117631

Qy         62 CGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAA 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     117630 CGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAA 117571

Qy        122 TTTCTCCCGTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTC 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     117570 TTTCTCCCGTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTC 117511

Qy        182 CGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCC 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     117510 CGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCC 117451

Qy        242 TCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAG 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     117450 TCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGGTAT 117391

Qy        302 GAATTTTCCAGCAAT 316
              |||  |||    |||
Db     117390 GAAACTTCTTCAAAT 117376


RESULT 6
AC019012/c
LOCUS       AC019012               130299 bp    DNA     linear   PLN 28-APR-2000
DEFINITION  Genomic Sequence For Arabidopsis thaliana Clone T3J11 From
            Chromosome V, complete sequence.
ACCESSION   AC019012
VERSION     AC019012.4  GI:7658323
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 130299)
  AUTHORS   Huang,E.N., Nascimento,L.U., de la Bastide,M., Preston,R.R.,
            Vil,D.M., Spiegel,L.A., See,L.H., Rodriguez,M.A., Shah,R.S.,
            Shekher,M., Kirchoff,K.A., Baker,J.P., O'Shaughnessy,A.,
            Dedhia,N.N. and McCombie,W.R.
  TITLE     Genomic Sequence For Arabidopsis thaliana Clone T3J11 From
            Chromosome V, complete sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 130299)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1999) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
REFERENCE   3  (bases 1 to 130299)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2000) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
COMMENT     On Apr 28, 2000 this sequence version replaced gi:7158061.
            The assembly overlaps from base 1 to base 41329 with F11P10
            (Accession number AC018660).  The overlap is from base 41329 to
            base 1 on F11P10.
FEATURES             Location/Qualifiers
     source          1. .130299
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="V"
                     /clone="T3J11"
                     /clone_lib="TAMU"
BASE COUNT    40445 a  25132 c  25452 g  39270 t
ORIGIN

  Query Match             14.4%;  Score 151.2;  DB 8;  Length 130299;
  Best Local Similarity   90.0%;  Pred. No. 3e-28;
  Matches  162;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy        296 AGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTAC 355
              | |||||||||||||| ||||||| ||||||||||||||||||||||||||||| |||
Db     105979 AACTAGGAATTTTCCAACAATGAGATTTGGTGGTAGGATTTTGTATAGCAAGACTACTAT 105920

Qy        356 TGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATC 415
              ||||||||||||||||| | |||||||| |||| |||||||| ||||||||||||||||
Db     105919 TGAGGTTGATAAGCGAGGATTGCAGCTTCTTAAGGTTCTTGAAACCAAGAGAGATGAATC 105860

Qy        416 TGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCT 475
              ||||| |||||||||||||| |||||||||||| ||||||  |||||||| ||| | |
Db     105859 TGGAAAAGCTTTTGTTGGCTTTGATATTGAGTGGAAACCAAGAATTAGAAAATGTGCTTT 105800


RESULT 7
AC002341
LOCUS       AC002341               82484 bp    DNA     linear   PLN 11-MAR-2002
DEFINITION  Arabidopsis thaliana chromosome 2 clone T14G11 map TEn5, complete
            sequence.
ACCESSION   AC002341
VERSION     AC002341.3  GI:20196903
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 82484)
  AUTHORS   Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,
            Spriggs,T.A., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
            Somerville,C.R. and Venter,J.C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 82484)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 82484)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598354.
FEATURES             Location/Qualifiers
     source          1. .82484
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="TEn5"
                     /clone="T14G11"
```

```
misc_feature    1. .11844
                /note="overlap with BAC clone T1B8
                (U78721:102301. .114144)."
gene            1141. .1212
                /gene="At2g33890"
tRNA            1141. .1212
                /gene="At2g33890"
                /product="tRNA-Pro"
gene            2114. .2185
                /gene="At2g33900"
tRNA            2114. .2185
                /gene="At2g33900"
                /product="tRNA-Pro"
gene            2946. .3017
                /gene="At2g33910"
tRNA            2946. .3017
                /gene="At2g33910"
                /product="tRNA-Pro"
gene            3898. .3969
                /gene="At2g33920"
tRNA            3898. .3969
                /gene="At2g33920"
                /product="tRNA-Pro"
gene            4411. .4482
                /gene="At2g33930"
tRNA            4411. .4482
                /gene="At2g33930"
                /product="tRNA-Pro"
gene            4853. .4924
                /gene="At2g33940"
tRNA            4853. .4924
                /gene="At2g33940"
                /product="tRNA-Pro"
gene            5892. .5963
                /gene="At2g33950"
tRNA            5892. .5963
                /gene="At2g33950"
                /product="tRNA-Pro"
gene            6810. .6881
                /gene="At2g33960"
tRNA            6810. .6881
                /gene="At2g33960"
                /product="tRNA-Pro"
gene            7276. .7347
                /gene="At2g33970"
tRNA            7276. .7347
                /gene="At2g33970"
                /product="tRNA-Pro"
gene            complement(7510. .10595)
                /gene="At2g33980"
                /note="T14G11.10; predicted by genefinder; supported by
                cDNA: gi_15293224_gb_AY051046 1_"
mRNA            complement(join(<7510. .7581,8935. .9025,9129. .9326,
                9417. .9536,9950. .10092,10280. .>10595))
                /gene="At2g33980"
CDS             complement(join(7541. .7581,8935. .9025,9129. .9326,
                9417. .9536,9950. .10092,10280. .10595))
                /gene="At2g33980"
                /codon_start=1
                /product="expressed protein"
                /protein_id="AAB67616.2"
                /db_xref="GI:20196908"
                /translation="MKSGASAASPTAKSFNFGSSRLLALAQQLRVYKPPLSSSFDERE
                ELLAYKESTRKSITHVGFQESMAPVRFRPKKAAVLICLFEGDDGDLRVILTKRSSTLS
                THSGEVSLPGGKAEEHDKDDGITATREAEEEIGLDPSLVDVVAFLEPFLSQHLLRVIP
                VVGILWDRKAFNPTPNPAEVEAVLDAPFEMFLKDENRRSEEFDWMGEKHLVHFFDYKT
                GDSDYVIWGLTARILIRAATVVYQRPPAFIEQKPNLKYSKMNQATRLYGWLKPLANLH
                ESCENGRLLGHVPAYCLRYLVGISSR"
gene            complement(11257. .12385)
                /gene="At2g33990"
                /note="T14G11.11"
mRNA            complement(join(<11257. .11481,11607. .11720,11815. .12042,
                12131. .12247,12320. .>12385))
                /gene="At2g33990"
CDS             complement(join(11257. .11481,11607. .11720,11815. .12042,
                12131. .12247,12320. .12385))
                /gene="At2g33990"
                /codon_start=1
                /product="putative SF16 protein (Helianthus annuus)"
                /protein_id="AAB67617.1"
                /db_xref="GI:2342719"
                /translation="MGSGNLIKAIIRLKKSKQGTEKKKTSAVKPKKGSKKKGTSLVTR
                SEDWAATRIQTAFKAYKARKSLRRLKGIARAKLSTEKQSVKNQAVVTLRYLHSWSKIQ
                SEIKARRVCMVTEWRLKNKRLEHQQKLEAKLHDVEVEWNGGSETKDEILERILQREEA
                TIKRERALAYAFSHQLGNTNWGWSWKERWISARPWEVRYSVTPKKPKSSKTDSNSPAK
                RTVSLSSVPAKTPFPGARNTVKPRRLSFPGA"
repeat_region   complement(13092. .13112)
                /rpt_family="AT_rich"
repeat_region   14529. .14597
                /rpt_family="(A)n"
repeat_region   14676. .14736
                /rpt_family="AT_rich"
gene            15932. .16387
                /gene="At2g34000"
                /note="T14G11.12; predicted by grail"
mRNA            <15932. .>16387
                /gene="At2g34000"
CDS             15932. .16387
                /gene="At2g34000"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAB67618.1"
                /db_xref="GI:2342720"
                /translation="MGFNDPSLNTIILWFASVTSLVTISVIFALLIICLLKRRRFDVS
                PETENENQGRREPRCQGLSASVIAAFPTFSYKPDNNDPESNNQEIECPVCLGLIPKNV
                VIKVLPNCMHMFDEECIGKWLESHATCPVCRRLAEPMTSNGDKVLERIV"
gene            19624. .20888
                /gene="At2g34010"
                /note="T14G11.13; predicted by genscan"
mRNA            join(<19624. .19940,20493. .20631,20709. .>20888)
                /gene="At2g34010"
CDS             join(19624. .19940,20493. .20631,20709. .20888)
                /gene="At2g34010"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAB67619.1"
                /db_xref="GI:2342721"
                /translation="MCSNNNTSSGSYGELQLHDESSGSCRKKQKKDKVRRRGPGVAEL
                EKIRLQEEYKPPLSSSPSLPNIDHHHHTLFAPASSVYDLVMTSPNFSFPEKLASTLPV
                FPISPDATQNRSMGITPVQETGTTTSNPIAIDSPTSIPRHYPRFIPLGLQYEQQQQKL
                KDLDETMQWRSKKPFYSFIPSGDPSNDDQEQRPCDLFGSAADHGIDLNLKL"
gene            complement(22342. .24458)
                /gene="At2g34020"
                /note="T14G11.14; similar to T14G11 15"
mRNA            complement(join(<22342. .22978,23078. .23217,23724. .23905,
                24029. .>24458))
                /gene="At2g34020"
CDS             complement(join(22342. .22978,23078. .23217,23724. .23905,
                24029. .24458))
                /gene="At2g34020"
                /codon_start=1
                /product="unknown protein"
                /protein_id="AAB67620.1"
                /db_xref="GI:2342722"
                /translation="MQTFSFKRRPNHSHFPSTLFFFFLSLTLFISGVSSRVLSPVPLD
                NSSLISDGVHGASDYKYLTLDPPKNVSKAACIHVYGFLPCADNIGGYVFQVFSFGCLL
                IIGDYFLSQGRSKLFVIFEVGFYGGIVFPLLTMFPRIALMISPGLAATHEGALMIVGN
                NVGVTIGHTIFALTMQWGACVIFGLTSPNSDPSIRRGSIKRTSSYFDTADQKKSLDHA
                KFELMSEVHKHLQSFSPRTLIRDGQLSKESLKSLFDKIDRNKDGKIQISELKDLTVEF
                GVFGKMKCDINEFASTLLAEFDKDKNGELDENEFEEGIMKLLNHYKFDNQESPRQNNT
                YIYRTPSDSVKSLSQGEEAGVLKLEMPKQTLVAKFLSMSTLRAVTKVIGGMLIVVFLA
                KPFMVNIGLLSVSAGVPSFYSVFAVIPLVRNLKNTLSAHFCRKKDKARIASEKFSEVS
                LLSTTLSLTTFY"
repeat_region   24544. .24589
                /rpt_family="(GAGAA)n"
gene            complement(26146. .28119)
```

```
                     /gene="At2g34030"
                     /note="T14G11.15; similar to T14G11 14"
     mRNA            complement(join(<26146. .26469,26569. .26815,26898. .27037,
                     27489. .27655,27726. .>28119))
                     /gene="At2g34030"
     CDS             complement(join(26146. .26469,26569. .26815,26898. .27037,
                     27489. .27655,27726. .28119))
                     /gene="At2g34030"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAB67621.1"
                     /db_xref="GI:2342723"
                     /translation="MAGVTLFFSLLSIALIISGVSSRVLISHVPLNNSILISDGIHDA
                     LNHEFLTLDPPKSLSRTACVHVYGFLPCADNVEGYIFQVFSFGSLLIIGDYFLSEGRS
                     KLFVIFEVGFYGGIIFPLLTMFPRIALMLSTGLSLSRDVASSFIDDNVGLTVGHTVFS
                     LTIQWGACVVFSITGPRSDQAVRRTTSYLDTANQEKSLDHARFELMSEVHKHLKRFSP
                     KHLIKDGELSKESLKSLFKKTDKNKDGKIQISELKDLTIELSNFGRMRYDINELAKAF
                     LEDFDGDNDGELEENEFEEGIARLLKQYKFNVEDQRENQTEENGVLKLEIKPKKTLVT
                     KLLSMETLIATTEVIVGILIVLFLAKPFMLNIQLLSISAGIPSFYIVFAMIPFARNLK

  Query Match             13.5%;  Score 142;  DB 8;  Length 82484;
  Best Local Similarity   73.6%;  Pred. No. 7.3e-26;
  Matches  209;  Conservative    0;  Mismatches   70;  Indels    5;  Gaps    2;

Qy        485 GGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCA 544
              | |||| |  |||||||||| | |||||| |||||||||| |   |   ||  | |||
Db      53823 GTTTGCAAGAGTCCAGATATATATAGATAATAATTATTGTTGTTGAAGTGATCATATTCT 53882

Qy        545 TTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGG 604
              |  || |  || |  | |       |  |  ||   |||   |||  |  | ||||||
Db      53883 TGTTGTTTGCCATGGATGCTAAATTAACATAGTTCTGGAT---ACATATAAATAGGTAGA 53939

Qy        605 TATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGA 664
              || ||||||||||||||||| |||||| |||||||||| |||||||||   || |||||
Db      53940 TAATGGAATTGATGGTGACTATGTGAAACTTTTCCATGAGTATGGAGTTTCCATTAAAGA 53999

Qy        665 TGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGA--TAAAAAATGGGGC 722
              ||| ||| |||| || |||||| |||||| ||||||||||||||||  | ||||||||||
Db      54000 TGTAGAGAATCTGTCGGATTTACCCAACCGAAAAATTGGTGGAGATTTGAAAAATGGGGT 54059

Qy        723 CTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAA 766
              |||||| |||||||||||||||||||||||||||| | |  ||
Db      54060 CTTGCCGCACTAACTGAGACACTTGTTTGCAAAGAGGTACCTAA 54103


RESULT 8
AX375189
LOCUS       AX375189               4299 bp    DNA     linear   PAT 01-MAR-2002
DEFINITION  Sequence 17 from Patent WO0210362.
ACCESSION   AX375189
VERSION     AX375189.1  GI:19169931
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Levin,J.Z., Phillips,K.L., Budziszewski,G.J., Meins,F.J. and
            Glazov,E.A.
  TITLE     Methods of controlling gene expression and gene silencing
  JOURNAL   Patent: WO 0210362-A 17 07-FEB-2002;
            Syngenta Participations AG (CH) ; Novartis Forschungsstiftung
            Zweigniederlassung (CH)
FEATURES             Location/Qualifiers
     source          1. .4299
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1400 a    759 c    950 g   1190 t
ORIGIN

  Query Match             10.6%;  Score 111;  DB 6;  Length 4299;
  Best Local Similarity   54.1%;  Pred. No. 8e-18;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || ||||||  ||||
Db      . 202 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 261

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||    || || |||||     | ||| | |||||   |
Db        262 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 318

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |       || | | | || ||  |  |  ||
Db        319 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 378

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||       |||
Db        379 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 438

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  |||||  | |||||
Db        439 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 498

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |      |      | |  ||| ||||  |   ||   | |    || | |
Db        499 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 558

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||     ||     ||| |   | | || |||       | ||||||  |
Db        559 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 618

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        619 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 678

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        679 TTAGAGATTTTGGATGATACTGT 701


RESULT 9
AX333649
LOCUS       AX333649               5189 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 4158 from Patent WO0194629.
ACCESSION   AX333649
VERSION     AX333649.1  GI:18124368
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 4158 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .5189
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1663 a    889 c   1129 g   1508 t
ORIGIN

  Query Match             10.6%;  Score 111;  DB 6;  Length 5189;
  Best Local Similarity   54.1%;  Pred. No. 8e-18;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || ||||||  ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||    || || |||||     | ||| | |||||   |
```

```
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572

Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632

Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692

Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752

Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812

Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872

Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895

Db        910 TTAGAGATTTTGGATGATACTGT 932
```

```
RESULT 10
E16596
LOCUS       E16596                  5189 bp    DNA     linear   PAT 28-JUL-1999
DEFINITION  Human WS mRNA, complete cds.
ACCESSION   E16596
VERSION     E16596.1  GI:5711279
KEYWORDS    JP 1998201498-A/1.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5189)
  AUTHORS   Matsumoto,T., Goto,M. and Furuichi,Y.
  TITLE     DETECTION OF MUTATION IN PATHOGENIC GENE OF HUMAN WERNER SYNDROME
  JOURNAL   Patent: JP 1998201498-A 1 04-AUG-1998;
            EIJIIN KENKYUSHO:KK
COMMENT     OS   Homo sapiens (human)
            PN   JP 1998201498-A/1
            PD   04-AUG-1998
            PF   24-JAN-1997 JP 1997011268
            PI   MATSUMOTO TAKEHISA, GOTO MAKOTO, FURUICHI YASUHIRO PC
            C12Q1/68,C07H21/04,C12N15/09,G01N33/50,G01N33/566; CC
            strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .5189
            FT                   /organism='Homo sapiens'
            FT   CDS             232. .4530
            FT                   /product='WS'
            FT                   /note='werner's syndrome'.
FEATURES             Location/Qualifiers
     source          1. .5189
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1663 a    889 c   1129 g   1508 t
ORIGIN

  Query Match             10.6%;  Score 111;  DB 6;  Length 5189;
  Best Local Similarity   54.1%;  Pred. No. 8e-18;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;
```

```
Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455

Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515

Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572

Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632

Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692

Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752

Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812

Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872

Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895

Db        910 TTAGAGATTTTGGATGATACTGT 932
```

```
RESULT 11
HUMDR
LOCUS       HUMDR                   5189 bp ss-DNA     linear   PRI 25-APR-1996
DEFINITION  Homo sapiens Werner syndrome gene, complete cds.
ACCESSION   L76937
VERSION     L76937.1  GI:1280207
KEYWORDS    Werner's syndrome; progeria.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5189)
  AUTHORS   Yu,C.-E., Oshima,J., Fu,Y.-H., Wijsman,E.M., Hisama,F., Alisch,R.,
            Matthews,S., Nakura,J., Miki,T., Ouais,S., Martin,G.M., Mulligan,J.
            and Schellenberg,G.D.
  TITLE     Positional cloning of the Werner's syndrome gene
  JOURNAL   Science 272 (5259), 258-262 (1996)
  MEDLINE   96181115
   PUBMED   8602509
FEATURES             Location/Qualifiers
     source          1. .5189
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     mRNA            1. .5189
                     /partial
     CDS             232. .4530
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="AAC41981.1"
                     /db_xref="GI:1280208"
                     /translation="MSEKKLETTAQQRKCPEWMNVQNKRCAVEERKACVRKSVFEDDL
                     PFLEFTGSIVYSYDASDCSFLSEDISMSLSDGDVVGFDMEWPPLYNRGKLGKVALIQL
                     CVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKLKNFVE
                     LTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLYAATDAY
                     AGFIIYRNLEILDDTVQRFAINKEEEILLSDMNKQLTSISEEVMDLAKHLPHAFSKLE
                     NPRRVSILLKDISENLYSLRRMIIGSTNIETELRPSNNLNLLSFEDSTTGGVQQKQIR
```

```
                     EHEVLIHVEDETWDPTLDHLAKHDGEDVLGNKVERKEDGFEDGVEDNKLKENMERACL
                     MSLDITEHELQILEQQSQEEYLSDIAYKSTEHLSPNDNENDTSYVIESDEDLEMEMLK
                     HLSPNDNENDTSYVIESDEDLEMEMLKSLENLNSGTVEPTHSKCLKMERNLGLPTKEE
                     EEDDENEANEGEEDDDKDFLWPAPNEEQVTCLKMYFGHSSFKPVQWKVIHSVLEERRD
                     NVAVMATGYGKSLCFQYPPVYVGKIGLVISPLISLMEDQVLQLKMSNIPACFLGSAQS
                     ENVLTDIKLGKYRIVYVTPEYCSGNMGLLQQLEADIGITLIAVDEAHCISEWGHDFRD
                     SFRKLGSLKTALPMVPIVALTATASSSIREDIVRCLNLRNPQITCTGFDRPNLYLEVR
                     RKTGNILQDLQPFLVKTSSHWEFEGPTIIYCPSRKMTQQVTGELRKLNLSCGTYHAGM
                     SFSTRKDIHHRFVRDEIQCVIATIAFGMGINKADIRQVIHYGAPKDMESYYQEIGRAG
                     RDGLQSSCHVLWAPADINLNRHLLTEIRNEKFRLYKLKMMAKMEKYLHSSRCRRQIIL
                     SHFEDKQVQKASLGIMGTEKCCDNCRSRLDHCYSMDDSEDTSWDFGPQAFKLLSAVDI
                     LGEKFGIGLPILFLRGSNSQRLADQYRRHSLFGTGKDQTESWWKAFSRQLITEGFLVE
                     VSRYNKFMKICALTKKGRNWLHKANTESQSLILQANEELCPKKFLLPSSKTVSSGTKE
                     HCYNQVPVELSTEKKSNLEKLYSYKPCDKISSGSNISKKSIMVQSPEKAYSSSQPVIS
                     AQEQETQIVLYGKLVEARQKHANKMDVPPAILATNKILVDMAKMRPTTVENVKRIDGV
                     SEGKAAMLAPLLEVIKHFCQTNSVQTDLFSSTKPQEEQKTSLVAKNKICTLSQSMAIT
                     YSLFQEKKMPLKSIAESRILPLMTIGMHLSQAVKAGCPLDLERAGLTPEVQKIIADVI
                     RNPPVNSDMSKISLIRMLVPENIDTYLIHMAIEILKHGPDSGLQPSCDVNKRRCFPGS
                     EEICSSSKRSKEEVGINTETSSAERKRRLPVWFAKGSDTSKKLMDKTKRGGLFS"
     polyA_signal    5159. .5164
     polyA_site      5189
BASE COUNT     1663 a    889 c   1129 g   1508 t
ORIGIN

  Query Match            10.6%;  Score 111;  DB 9;  Length 5189;
  Best Local Similarity   54.1%;  Pred. No. 8e-18;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || |||||  ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||    || || |||||     | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  ||||| | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |     |     | |  ||| ||||  |   ||   | |    ||  |  |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||     ||    ||| |   | | || |||      | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932


RESULT 12
AF091214
LOCUS       AF091214                5208 bp    mRNA    linear   PRI 08-OCT-1998
DEFINITION  Homo sapiens WRN (WRN) mRNA, complete cds.
ACCESSION   AF091214
VERSION     AF091214.1  GI:3719420
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5208)
  AUTHORS   Paeper,B.W., Gayle,M., Brady,W., Swartz,A., Gillett,L.A.,
            Alisch,R.S., Mulligan,J., Galas,D. and Fu,Y.-H.
  TITLE     Genomic structure of the human Werner's gene and cloning of the
            mouse homolog
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5208)
  AUTHORS   Paeper,B.W., Gayle,M., Brady,W., Swartz,A., Gillett,L.A.,
            Alisch,R.S., Mulligan,J., Galas,D. and Fu,Y.-H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-1998) Genomics, Chiroscience R&D Inc., 1631 220th
            Street SE, Bothell, WA 98021, USA
FEATURES             Location/Qualifiers
     source          1. .5208
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .5208
                     /gene="WRN"
     5'UTR           1. .231
                     /gene="WRN"
     CDS             232. .4530
                     /gene="WRN"
                     /note="Werner syndrome"
                     /codon_start=1
                     /product="WRN"
                     /protein_id="AAC63361.1"
                     /db_xref="GI:3719421"
                     /translation="MSEKKLETTAQQRKCPEWMNVQNKRCAVEERKACVRKSVFEDDL
                     PFLEFTGSIVYSYDASDCSFLSEDISMSLSDGDVVGFDMEWPPLYNRGKLGKVALIQL
                     CVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKLKNFVE
                     LTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLYAATDAY
                     AGFIIYRNLEILDDTVQRFAINKEEEILLSDMNKQLTSISEEVMDLAKHLPHAFSKLE
                     NPRRVSILLKDISENLYSLRRMIIGSTNIETELRPSNNLNLLSFEDSTTGGVQQKQIR
                     EHEVLIHVEDETWDPTLDHLAKHDGEDVLGNKVERKEDGFEDGVEDNKLKENMERACL
                     MSLDITEHELQILEQQSQEEYLSDIAYKSTEHLSPNDNENDTSYVIESDEDLEMEMLK
                     HLSPNDNENDTSYVIESDEDLEMEMLKSLENLNSGTVEPTHSKCLKMERNLGLPTKEE
                     EEDDENEANEGEEDDDKDFLWPAPNEEQVTCLKMYFGHSSFKPVQWKVIHSVLEERRD
                     NVAVMATGYGKSLCFQYPPVYVGKIGLVISPLISLMEDQVLQLKMSNIPACFLGSAQS
                     ENVLTDIKLGKYRIVYVTPEYCSGNMGLLQQLEADIGITLIAVDEAHCISEWGHDFRD
                     SFRKLGSLKTALPMVPIVALTATASSSIREDIVRCLNLRNPQITCTGFDRPNLYLEVR
                     RKTGNILQDLQPFLVKTSSHWEFEGPTIIYCPSRKMTQQVTGELRKLNLSCGTYHAGM
                     SFSTRKDIHHRFVRDEIQCVIATIAFGMGINKADIRQVIHYGAPKDMESYYQEIGRAG
                     RDGLQSSCHVLWAPADINLNRHLLTEIRNEKFRLYKLKMMAKMEKYLHSSRCRRQIIL
                     SHFEDKQVQKASLGIMGTEKCCDNCRSRLDHCYSMDDSEDTSWDFGPQAFKLLSAVDI
                     LGEKFGIGLPILFLRGSNSQRLADQYRRHSLFGTGKDQTESWWKAFSRQLITEGFLVE
                     VSRYNKFMKICALTKKGRNWLHKANTESQSLILQANEELCPKKFLLPSSKTVSSGTKE
                     HCYNQVPVELSTEKKSNLEKLYSYKPCDKISSGSNISKKSIMVQSPEKAYSSSQPVIS
                     AQEQETQIVLYGKLVEARQKHANKMDVPPAILATNKILVDMAKMRPTTVENVKRIDGV
                     SEGKAAMLAPLLEVIKHFCQTNSVQTDLFSSTKPQEEQKTSLVAKNKICTLSQSMAIT
                     YSLFQEKKMPLKSIAESRILPLMTIGMHLSQAVKAGCPLDLERAGLTPEVQKIIADVI
                     RNPPVNSDMSKISLIRMLVPENIDTYLIHMAIEILKHGPDSGLQPSCDVNKRRCFPGS
                     EEICSSSKRSKEEVGINTETSSAERKRRLPVWFAKGSDTSKKLMDKTKRGGLFS"
     3'UTR           4531. .5208
                     /gene="WRN"
BASE COUNT     1681 a    887 c   1131 g   1509 t
ORIGIN

  Query Match            10.6%;  Score 111;  DB 9;  Length 5208;
  Best Local Similarity   54.1%;  Pred. No. 8e-18;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || |||||  ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||    || || |||||     | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
```

```
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632

Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692

Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752

Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812

Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872

Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895

Db        910 TTAGAGATTTTGGATGATACTGT 932
```

```
RESULT 13
AF067418
LOCUS       AF067418               5143 bp    mRNA    linear   VRT 09-OCT-1998
DEFINITION  Xenopus laevis focus forming activity 1 (FFA-1) mRNA, complete cds.
ACCESSION   AF067418
VERSION     AF067418.1  GI:3420290
KEYWORDS    .
SOURCE      Xenopus laevis.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 5143)
  AUTHORS   Yan,H., Chen,C.Y., Kobayashi,R. and Newport,J.
  TITLE     Replication focus-forming activity 1 and the Werner syndrome gene
            product
  JOURNAL   Nat. Genet. 19 (4), 375-378 (1998)
  MEDLINE   98361165
   PUBMED   9697700
REFERENCE   2  (bases 1 to 5143)
  AUTHORS   Yan,H., Chen,C.Y., Kobayashi,R. and Newport,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-1998) Fox Chase Cancer Center, 7701 Burholme
            Avenue, Philadelphia, PA 19111, USA
FEATURES             Location/Qualifiers
     source          1. .5143
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /cell_type="oocyte"
     gene            1. .5143
                     /gene="FFA-1"
     CDS             160. .4470
                     /gene="FFA-1"
                     /function="DNA helicase"
                     /codon_start=1
                     /product="focus forming activity 1"
                     /protein_id="AAC63512.1"
                     /db_xref="GI:3420291"
                     /translation="MTSLQRKLPEWMSVKQQEDRIDDAKKSFCKKNILEDNLPFMKFN
                     GSIVYSYESNDCSLLSEDIRSSLLEEDVLGFDIEWPPVYTKGKTGKVALIQVCVSEKK
                     CYLFHISPMAGFPKGLKRLLEDESVRKVGVGIEGDQWKLMSDYELKLKGFIELSEMAN
                     QKLRCKEKWTFNGLIKHLFKEQLYKRKSYRCSNWDIFLLTEDQKLYAATDAYAGLLIY
                     KKLEGMDAHESDSFRVGREGVADCKGVKRQLTDLSKGLMDLVNQVPNSFGCYTEAVRA
                     VDILEDLSEKLEELRNIMKEASKAEGNGLHFQNSEDCSKKDKSILHVACKESLAEHKM
                     DCKNADSQNNKDIDSCQNENRDEDFFMTLGISEEELYMMEREDDKKQTNPDYKLNKDS
```

```
                     CDTNEEKDMSYVIESDEDFDSEIIKSLEDLDNSTEEALGTGVPQAGLIPAKSVDTVAD
                     EEEDEGIEEEDDDDDWDPSMPEPSAQHISCLKTYFGHSSFKPVQWKVVHSVLRERRDN
                     LVVMATGYGKSLCYQFAPVYTSGIGIVICPLISLMEDQVLQLEMSNISSCFLGSAQSK
                     NVLQDVKDGKMRVIYMTPEFCSRGISLLQDLDNRYGITLIAIDEAHCISEWGHDFRSA
                     YRSLGSLKRMLPNVPIVALTATASPSIREDITKSLNLHNPQVTCTSFDRPNLYLDVAR
                     KTTNISIDLRQFLIKKQQGSGWEFEGATIVYCPTRKTSEQVTAELIKLGIACGTYHAG
                     MGIKQRREVHHRFMRDEIHCVVATVAFGMGINKPDIRKVIHYGAPKEMESYYQEIGRA
                     GRDGLPSCCHALWAQADMNFNRHMLGEIPNKGFREYKLKMLTKMEKYLNSSTCRRKII
                     LSHFEDKQLRKASSGIMGTEKCCDNCKTRLICNISINDTEDNLQDFGPQAYKFISAVD
                     VLGQKFGTGVPVLFLRGSTSQRVPDRFRNHSLFSSGKDQTEAFWKVLARQLITEGYLQ
                     ESSGQTKFSTICGLTSKGSNWLIKANNEQCPSLLLPSNNELCLQRTRVSNFSSAQAHS
                     SMVPHASSNTRSSMPKAGPEKMELKDKFSYQEAERLSKAAGVSKSSFKLQTPCKLSRP
                     PEPEVSPRERELQTTLYGRLVVARQKIASERDILPAVLATNKVLVDMAKLRPTTSENM
                     KKLDGVSEAKSAMLAPLLEVVKEFCIANSLKVDVFSGSVSQSESTFFTPREQERISLP
                     ESQRMSYSLFQEQNLSLKKIADVRCLSMAVVGMHLWQALKAGYSFDVQRAGLTPEMKK
                     LITYAIKKPPINSDLSSFKAIREYVPANIDGYPIRMVISLLEKEGSSGAQGQPEFPTQ
                     KTLIQTEENPKNVSVQNTKHKVTMGKSMWIEKKPTQPATAELEVTKGKALAPIMLASW
                     NEASLDADTEELFSESQSSTTRPRRRLPEWFGSTKGNAATRCIQESKNLGEEKGSFFD
                     "
BASE COUNT     1689 a    951 c   1104 g   1399 t
ORIGIN

  Query Match           10.5%;  Score 110.6;  DB 5;  Length 5143;
  Best Local Similarity   54.1%;  Pred. No. 1e-17;
  Matches  271;  Conservative    0;  Mismatches  224;  Indels    6;  Gaps    2;

Qy        423 GCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGG 482

Db        370 GATGTACTTGGTTTTGATATAGAATGGCCTCCTGTTTATACAAAGGG---AAAAACTGGA 426

Qy        483 AAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTT 542

Db        427 AAAGTTGCTCTTATACAAGTGTGTGTGTCTGAGAAAAAATGCTATTTGTTTCATATTTCA 486

Qy        543 CAT---TCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAG 599

Db        487 CCTATGGCTGGCTTTCCTAAGGGTTTGAAGAGGCTGCTAGAAGATGAGTCTGTTAGAAAA 546

Qy        600 GTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATC 659

Db        547 GTTGGTGTGGGAATTGAAGGAGACCAGTGGAAGTTAATGAGTGACTATGAGCTGAAACTT 606

Qy        660 AAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGG 719

Db        607 AAAGGCTTTATTGAACTTTCGGAAATGGCTAACCAAAAACTACGGTGCAAGGAAAAGTGG 666

Qy        720 GGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATC 779

Db        667 ACCTTTAATGGTTTGATCAAACATCTTTTCAAGGAGCAACTTTATAAAAGAAAGTCTTAC 726

Qy        780 AGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACG 839

Db        727 CGTTGCAGCAACTGGGATATTTTTCTTCTCACTGAAGACCAGAAATTGTATGCTGCCACT 786

Qy        840 GATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGT 899

Db        787 GATGCTTATGCTGGGCTACTCATTTACAAGAAACTAGAGGGCATGGATGCTCACGAGAGT 846

Qy        900 GGCTCATAACGTGAAGGAGGA 920

Db        847 GATTCCTTCCGTGTTGGCAGA 867
```

```
RESULT 14
ATFCA0/c
LOCUS       ATFCA0                200576 bp    DNA     linear   PLN 28-JUN-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
            No. 0.
ACCESSION   Z97335
VERSION     Z97335.2  GI:5280985
KEYWORDS    .
SOURCE      Arabidopsis thaliana.
  ORGANISM  Arabidopsis thaliana
```

```
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 200576)
  AUTHORS   Bevan,M., Stiekema,W., Murphy,G., Wambutt,R., Pohl,T., Terryn,N.,
            Kreis,M., Kavanagh,T., Entian,K.D., Rieger,M., James,R.,
            Puigdomenech,P., Hatzopoulos,P., Obermaier,B., Duesterhoft,A.,
            Jones,J., Palme,K., Ansorge,W., Delseny,M., Bancroft,I.,
            Mewes,H.W., Schueller,C. and Chalwatzis,N.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200576)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     On Jun 30, 1999 this sequence version replaced gi:2244747.
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATFCA1 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .200576
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     source          1. .63289
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="BAC TAMU20K8"
     source          62805. .140141
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="BAC TAMU6G17"
     source          134184. .200576
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="BAC TAMU8H22"
     gene            1753. .4428
                     /gene="dl3000w"
     CDS             join(1753. .3194,3261. .4428)
                     /gene="dl3000w"
                     /note="similarity to Cf-2.2, Solanum pimpinellifolium"
                     /codon_start=1
                     /product="disease resistance Cf-2 like protein"
                     /protein_id="CAB10171.1"
                     /db_xref="GI:2244748"
                     /db_xref="SPTREMBL:O23253"
                     /translation="MITIIWSLCLIFCLSNSILVIAKDLCLPDQRDALLEFKNEFSIP
                     SPDSDLMLILQTTAKWRNNTDCCSWGGISCDPKTGVVVELDLGNSDLNGRLRSNSSLF
                     RLQHLQSLDLSYNDLSCTLPDSSGNFKYLRVLNLLGCNLFGEIPTSLRSLSYLTDLDL
                     SYNDDLTGEILDSMGNLKHLRVLSLTSCKFTGKIPSSLGNLTYLTDLDLSWNYFTGEL
                     PDSMGNLKSLRVLNLHRCNFFGKIPTSLGSLSNLTDLDISKNEFTSEGPDSMSSLNRL
                     TDFQLMLLNLSSLTNVDLSSNQFKAMLPSNMSSLSKLEAFDISGNSFSGTIPSSLFML
                     PSLIKLDLGTNDFSGPLKIGNISSPSNLQELYIGENNINGPIPRSILKLVGLSALSLS
                     FWDTGGIVDFSIFLQLKSLRSLDLSGINLNISSSHHLPSHMMHLILSSCNISQFPKFL
                     ENQTSLYHLDISANQIEGQVPEWLWRLPTLSFIASDNKFSGEIPRAVCEIGTLVLSNN
                     NFSGSIPPCFEISNKTLSILHLRNNSLSGVIPEESLHGYLRSLDVGSNRLSGQFPKSL
                     INCSYLQFLNVEENRINDTFPSWLKSLPNLQLLVLRSNEFHGPIFSPGDSLSFSKLRF
                     FDISENRFSGVLPSDYFVGWSVMSSFVDIIDNTPGFTVVGDDQESFHKSVVLTIKGLN
                     MELVGSGFEIYKTIDVSGNRLEGDIPESIGILKELIVLNMSNNAFTGHIPPSLSNLSN
                     LQSLDLSQNRLSGSIPGELGELTFLARMNFSYNMLEGPIPQGTQIQSQNSSSFAENPG
                     LCGAPLQKKCGGEEEEDKEKEEKDKGLSWVAAAIGYVPGLFCGLAIGHILTSYKRDWF
                     MRIFSCFSSPLKK"
     exon            1753. .3194
                     /gene="dl3000w"
                     /number=1
     intron          3195. .3260
                     /gene="dl3000w"
                     /number=1
     exon            3261. .4428
                     /gene="dl3000w"
                     /number=2
     gene            5905. .7913
                     /gene="dl3005c"
     gene            complement(join(5905. .6149,6243. .6344,6436. .6826,
                     7236. .7913))
                     /gene="dl3005c"
     CDS             complement(join(5905. .6149,6243. .6344,6436. .6826,
                     7236. .7913))
                     /gene="dl3005c"
                     /note="strong similarity to glycine
                     hydroxymethyltransferase (EC 2.1.2.1)
                     contains EST gb:1565775"
                     /codon_start=1
                     /product="hydroxymethyltransferase"
                     /protein_id="CAB10172.1"
                     /db_xref="GI:2244749"
                     /db_xref="SPTREMBL:O23254"
                     /translation="MEPVSSWGNTSLVSVDPEIHDLIEKEKRRQCRGIELIASENFTS
                     FAVIEALGSALTNKYSEGIPGNRYYGGNEFIDEIENLCRSRALEAFHCDPAAWGVNVQ
                     PYSGSPANFAAYTALLQPHDRIMGLDLPSGGHLTHGYYTSGGKKISATSIYFESLPYK
                     VNFTTGYIDYDKLEEKALDFRPKLLICGGSAYPRDWDYARFRAIADKVGALLLCDMAH
                     ISGLVAAQEAANPFEYCDVVTTTTHKSLRGPRAGMIFYRKGPKPPKKGQPEGAVYDFE
                     DKINFAVFPALQGGPHNHQIGALAVALKQANTPGFKVYAKQVKANAVALGNYLMSKGY
                     QIVTNGTENHLVLWDLRPLGLTGNKVEKLCDLCSITLNKNAVFGDSSALAPGGVRIGA
                     PAMTSRGLVEKDFEQIGEFLSRAVTLTLDIQKTYGKLLKDFNKGLVNNKDLDQLKADV
                     EKFSASYEMPGFLMSEMKYKD"
     exon            complement(5905. .6149)
                     /gene="dl3005c"
                     /number=1
     intron          complement(6150. .6242)
                     /gene="dl3005c"
                     /number=1
     exon            complement(6243. .6344)
                     /gene="dl3005c"
                     /number=2
     intron          complement(6345. .6435)
                     /gene="dl3005c"
                     /number=2
     exon            complement(6436. .6826)
                     /gene="dl3005c"
                     /number=3
     intron          complement(6827. .7235)
                     /gene="dl3005c"
                     /number=3
     exon            complement(7236. .7913)
                     /gene="dl3005c"
                     /number=4
     gene            12823. .14568
                     /gene="dl3010w"
     CDS             join(12823. .13533,13822. .14568)
                     /gene="dl3010w"
                     /note="strong similarity to adenosylhomocysteinasE (EC
                     3.3.1.1)"
                     /codon_start=1
                     /product="adenosylhomocysteinase"
                     /protein_id="CAB10173.1"
                     /db_xref="GI:2244750"
                     /db_xref="SWISS-PROT:O23255"
                     /translation="MALLVEKTSSGREYKVKDMSQADFGRLELELAEVEMPGLMACRT
                     EFGPSQPFKGARITGSLHMTIQTAVLIETLTALGAEVRWCSCNIFSTQDHAAAAIARD
                     SAAVFAWKGETLQEYWWCTERALDWGPGGGPDLIVDDGGDATLLIHEGVKAEEIFEKT
                     GQVPDPTSTDNPEFQIVLSIIKEGLQVDPKKYHKMKERLVGVSEETTTGVKRLYQMQQ
                     NGTLLFPAINVNDSVTKSKFDNLYGCRHSLPDGLMRATDVMIAGKVAVICGYGDVGKG
                     CAAAMKTAGARVIVTEIDPICALQALMEGLQVLTLEDVVSEADIFVTTTGNKDIIMVD
                     HMRKMKNNAIVCNIGHFDNEIDMLGLETYPGVKRITIKPQTDRWVFPETKAGIIVLAE
                     GRLMNLGCATGHPSFVMSCSFTNQVIAQLELWNEKASGKYEKKVYVLPKHLDEKVALL
                     HLGKLGARLTKLSKDQSDYVSIPIEGPYKPPHYRY"
     exon            12823. .13533
                     /gene="dl3010w"
                     /number=1
```

```
     intron          13534. .13821
                     /gene="dl3010w"
                     /number=1
     exon            13822. .14568
                     /gene="dl3010w"
                     /number=2
     gene            16093. .16501
                     /gene="dl3015c"
     gene            complement(join(16093. .16270,16317. .16501))
                     /gene="dl3015c"
     CDS             complement(join(16093. .16270,16317. .16501))
                     /gene="dl3015c"
                     /note="weak similarity to Walleye dermal sarcoma virus gag
                     polyprotein"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB10174.1"
                     /db_xref="GI:2244751"
                     /db_xref="SPTREMBL:O23256"
                     /translation="MFNSTALVIFAILFLLISADAFPIPSPNGEIDAMLIRNSIIGED
                     EDLMPTEISRRVLMAQKRNQELLTTIVALDKLILTAEAVTPLPDALETQTTSTHEGKI
                     QEIAKIGSFLAEASKLYD"
     exon            complement(16093. .16270)
                     /gene="dl3015c"
                     /number=1
     intron          complement(16271. .16316)
                     /gene="dl3015c"
                     /number=1
     exon            complement(16317. .16501)
                     /gene="dl3015c"
                     /number=2
     gene            18573. .20123
                     /gene="dl3020c"
     gene            complement(join(18573. .18914,18997. .19137,19212. .19554,
                     19603. .19719,19762. .20123))
                     /gene="dl3020c"
     CDS             complement(join(18573. .18914,18997. .19137,19212. .19554,
                     19603. .19719,19762. .20123))
                     /gene="dl3020c"
                     /note="weak similarity to fusion protein 1a/1b - sugar
                     beet"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB10175.1"
                     /db_xref="GI:2244752"
                     /db_xref="SPTREMBL:O23257"

  Query Match             9.3%;  Score 97.8;  DB 8;  Length 200576;
  Best Local Similarity   76.9%;  Pred. No. 1.7e-14;
  Matches  133;  Conservative    0;  Mismatches   37;  Indels    3;  Gaps    1;

Qy        590 ACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGG 649
              | || | |     ||| || |||||||  |||||| ||| ||||||||||||||| ||||
Db        581 AGTTCTTATATGTGGTGTTTGAATTGACTGTGACTTTGTTAAGCTTTTCCATGACAATGG 522

Qy        650 AGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGA 709
               ||     |||||||||| || ||||||| |  ||||||| || ||||  ||| |  ||
Db        521 TGTATCCATCAAAGATGTCGAAGATCTTTTACGTTTAGCCCACAAAAACCTTGATTCAG- 463

Qy        710 TAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCC 762
               |||||||||||| |||||| ||||||||||||||||| ||||||||| | |
Db        462 --AAAAATGGGGCCGTGCCTCGCTAACTGAGACACTTGATTGCAAAGAGGTAC 412


RESULT 15
AR149285
LOCUS       AR149285               6476 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 5 from patent US 6228583.
ACCESSION   AR149285
VERSION     AR149285.1  GI:15113876
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6476)
  AUTHORS   Guarente,L.P. and Sinclair,D.A.
  TITLE     Assays for compounds which extend life span
  JOURNAL   Patent: US 6228583-A 5 08-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .6476
                     /organism="unknown"
BASE COUNT     2019 a   1300 c   1389 g   1768 t
ORIGIN

  Query Match             7.9%;  Score 82.8;  DB 6;  Length 6476;
  Best Local Similarity   53.2%;  Pred. No. 1.5e-10;
  Matches  199;  Conservative    0;  Mismatches  172;  Indels    3;  Gaps    1;

Qy        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540
              | |  || ||     ||||| | |||||   |   ||  | ||| |  |  | || ||||
Db        494 GCAGAGTCGCAGTGATCCAGTTGTGTGTGTCTGAGAACAAATGTTACTTGTTTCACATTT 553

Qy        541 TT---CATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAA 597
               |      || | | |||| ||  |  |  ||    |  | ||| |   | || ||   |
Db        554 CTTCCATGTCAGTTTTCCCCCAGGGATTAAAAATGTTACTAGAAAACAAATCAATTAAGA 613

Qy        598 AGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTA 657
              ||| |||  |||| ||||| || |||     ||| ||| | | ||| | ||  || |
Db        614 AGGCAGGGGTTGGGATTGAAGGGGACCAGTGGAAACTTCTGCGTGATTTTGACGTCAAGT 673

Qy        658 TCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAAT 717
              |  |   | ||| ||| ||  | ||| | |||||  ||||  |    |  |  | | |
Db        674 TGGAGAGTTTTGTGGAGCTGACGGATGTTGCCAATGAAAAGTTGAAGTGCGCAGAGACCT 733

Qy        718 GGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAA 777
              || ||||       ||   | |     | |  | ||| | ||  ||||    ||    |
Db        734 GGAGCCTCAATGGTCTGGTTAAACACGTCTTAGGGAAACAACTTTTGAAAGACAAGTCCA 793

Qy        778 TCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAA 837
              || |     | || |||     ||  || ||  |  || | |||  ||  || ||||| |
Db        794 TCCGCTGCAGCAATTGGAGTAATTTCCCCCTCACTGAGGACCAGAAACTGTATGCAGCCA 853

Qy        838 CGGATGCTTATGCT 851
              | ||||||||||||
Db        854 CTGATGCTTATGCT 867


Search completed: June 14, 2003, 07:52:41
Job time : 2954 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: June 14, 2003, 05:25:33 ; Search time 315 Seconds
(without alignments)
7499.512 Million cell updates/sec

Title: US-09-896-186B-23
Perfect score: 1049
Sequence: 1 accaaagcattaatttttat.....tttaaccgtccagaaactag 1049

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1049	100.0	1049	24	ABA91805	Arabidopsis RNase
	2	1039	99.0	1114	21	AAC51089	Arabidopsis thalia
	3	1002.2	95.5	1129	21	AAC40589	Arabidopsis thalia
	4	831	79.2	942	24	ABA91795	Arabidopsis RNase
	5	111	10.6	4299	24	ABA91803	Human Werner syndr
	6	111	10.6	5189	19	AAV65701	DNA fragment of a
	7	111	10.6	5189	24	ABL65821	Lung cancer relate
	8	111	10.6	5208	18	AAX83001	Human WRN coding s
	9	82.8	7.9	6476	20	AAX24302	Murine mWRN gene (
	10	81.2	7.7	1164	19	AAV35131	Mouse WRN helicase
	11	81.2	7.7	4206	19	AAV35115	Mouse WRN helicase
	12	81.2	7.7	4792	18	AAX83004	Mouse WRN coding s
	13	81.2	7.7	5058	19	AAV35114	Mouse WRN helicase
	14	51.2	4.9	29604	18	AAX83005	Partial mouse WRN
	15	44.4	4.2	449	22	ABA08508	Human secreted pro
	16	44.4	4.2	814	22	AAH06496	Human cDNA clone (
	17	44.4	4.2	2833	22	AAH14483	Human cDNA sequenc
	18	44.4	4.2	3149	20	AAX90441	Human secreted pro
	19	44.4	4.2	3149	22	AAS59270	Human cDNA encodin
	20	44.4	4.2	3149	24	ABA90939	Human polynucleoti
	21	42.4	4.0	651	24	ABN70543	Streptococcus poly
	22	42.4	4.0	690	24	ABN69471	Streptococcus poly
	23	42.4	4.0	2155561	24	ABN71527	Streptococcus poly
	24	40.8	3.9	13059	18	AAV74531	Staphylococcus aur
	25	40.4	3.9	15674	24	ABL70513	Chemically treated
	26	40.4	3.9	15674	24	ABL32362	Human immune syste
	27	40.4	3.9	15674	24	ABL34476	Human metastasis a
	28	40	3.8	580073	18	AAT58840	Mycoplasma genital
	29	39.2	3.7	1499	21	AAZ37308	Arabidopsis Abi4 c
	30	39.2	3.7	1500	21	AAZ37307	Arabidopsis Abi4 c
	31	39.2	3.7	1500	22	AAD05663	Arabidopsis thalia
c	32	38.2	3.6	439	22	ABA49878	Human breast cell
c	33	38.2	3.6	439	22	ABA67796	Human foetal liver
c	34	38.2	3.6	439	22	ABA34854	Probe #13320 for g
c	35	38.2	3.6	439	22	AAK16207	Human brain expres
c	36	38.2	3.6	439	22	AAK41953	Human bone marrow
c	37	38.2	3.6	439	22	AAI22718	Probe #12651 for g
c	38	38.2	3.6	439	22	AAI48016	Probe #16702 used
c	39	38.2	3.6	439	22	AAI08383	Probe #8374 used t
c	40	38.2	3.6	439	24	ABS15980	Human genome-deriv
c	41	38	3.6	3595	24	ABK84215	Human cDNA differe
	42	38	3.6	17534	24	ABK40025	Human chemically p
	43	37.4	3.6	681	22	AAH03262	Human cDNA clone (
	44	37.2	3.5	5433	21	AAA99265	Plasmodium yoelii
	45	37	3.5	8103	24	ABL33661	Human immune syste

ALIGNMENTS

RESULT 1
ABA91805
ID ABA91805 standard; cDNA; 1049 BP.
XX
AC ABA91805;
XX
DT 15-MAY-2002 (first entry)
XX
DE Arabidopsis RNase D related domain (3'-5' exonuclease domain) cDNA.
XX
KW RNase D; enzyme; 3'-5' exonuclease; gene expression; plant;
KW gene silencing; transgenic plant; gene; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 42..908
FT /*tag= a
FT /product= "3'-5'_exonuclease_domain"
XX
PN WO200210362-A2.
XX
PD 07-FEB-2002.
XX
PF 30-JUL-2001; 2001WO-EP08825.
XX
PR 01-AUG-2000; 2000US-222202P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
PA (NOVS) NOVARTIS FORSCHUNGSSTIFTUNG ZWEIGNIEDERL.
XX
PI Levin JZ, Phillips KL, Budziszewski GJ, Meins F, Glazov EA;

```
XX
DR   WPI; 2002-217119/27.
DR   P-PSDB; AAM50937.
XX
PT   Regulating gene expression in plants for controlling gene silencing,
PT   comprises altering the transcription or translation of an endonuclease
PT   nucleotide sequence encoding a polypeptide comprising an exonuclease
PT   domain -
XX
PS   Claim 4; Page 93-94; 102pp; English.
XX
CC   The present nucleotide sequence comprises cDNA encoding an
CC   Arabidopsis thaliana 3'-5' exonuclease domain (see AAM50937),
CC   specifically an RNase D related domain. The cDNA was identified
CC   on the basis of homology to the 3'-5' exonuclease sequence given in
CC   ABA91795 via 3'-and 5' RACE and TA-cloning. The sequence is nearly
CC   identical to an Arabidopsis exonuclease termed wrnexo. Nucleotide
CC   (see ABA91795-807) and polypeptide (see AAM50927-39) sequences for
CC   3'-5' exonuclease domains were identified using different screening
CC   methods and search algorithms. The invention encompasses the
CC   suppression or increase of gene silencing in plants. This is
CC   achieved by altering the expression in the plant cell of a
CC   nucleotide sequence encoding a polypeptide comprising an exonuclease
CC   domain, preferably a 3'-5' exonuclease domain, and especially an
CC   RNase D-related domain. Expression of the nucleotide sequence is
CC   altered by altering its transcription or translation. It is
CC   reduced e.g. by sense suppression, antisense suppression, homologous
CC   recombination, the use of ribozymes, dominant-negative mutants,
CC   aptamers, zinc finger proteins, double-stranded RNA, or insertional,
CC   point or deletion mutagenesis, and increased by overexpression.
CC   The methods are useful for altering or stabilizing the expression of
CC   a nucleotide sequence of interest in a plant cell or plant.
CC   Suppressing or decreasing expression of the nucleic acid molecule
CC   results in decreased levels of post transcriptional gene silencing
CC   and improved expression of genes of interest.
XX
SQ   Sequence 1049 BP; 292 A; 216 C; 220 G; 321 T; 0 other;

  Query Match           100.0%;  Score 1049;  DB 24;  Length 1049;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-297;
  Matches 1049;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ACCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGA 60

Qy         61 TCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACA 120

Qy        121 ATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCT 180

Qy        181 CCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGC 240

Qy        241 CTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA 300

Qy        301 GGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGG 360

Qy        361 TTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAA 420

Qy        421 TAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGG 480

Qy        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540

Qy        541 TTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGG 600

Qy        601 TAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCA 660

Qy        661 AAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGG 720

Qy        721 GCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCA 780

Qy        781 GGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGG 840

Qy        841 ATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTG 900

Qy        901 GCTCATAACGTGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCTCATAACGTGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAA 960

Qy        961 ATGATATGATACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATATGATACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGT 1020

Qy       1021 TCTGGTCCCTTTAACCGTCCAGAAACTAG 1049
              |||||||||||||||||||||||||||||
Db       1021 TCTGGTCCCTTTAACCGTCCAGAAACTAG 1049


RESULT 2
AAC51089
ID   AAC51089 standard; DNA; 1114 BP.
XX
AC   AAC51089;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 67235.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
```

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

```
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match           99.0%;  Score 1039;  DB 21;  Length 1114;
  Best Local Similarity  100.0%;  Pred. No. 2.4e-294;
  Matches 1039;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 TAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATCGACGACGC 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATCGACGACGC 60

Qy         71 TTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCG 120

Qy        131 TTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGG 180

Qy        191 CCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCAT 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCAT 240

Qy        251 CACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATTTTCC 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATTTTCC 300

Qy        311 AGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCG 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCG 360

Qy        371 AGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGT 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGT 420

Qy        431 TGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGC 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGC 480

Qy        491 GACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGG 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGG 540

Qy        551 TATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGG 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGG 600

Qy        611 AATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGA 660

Qy        671 GGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTC 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTC 720

Qy        731 ACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAA 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAA 780

Qy        791 CTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGC 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGC 840

Qy        851 TTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTGGCTCATAACG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTGGCTCATAACG 900

Qy        911 TGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAAATGATATGAT 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAAATGATATGAT 960

Qy        971 ACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGTTCTGGTCCCT 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGTTCTGGTCCCT 1020

Qy       1031 TTAACCGTCCAGAAACTAG 1049
              ||||||||||||||||||||
Db       1021 TTAACCGTCCAGAAACTAG 1039


RESULT 3
AAC40589
ID   AAC40589 standard; DNA; 1129 BP.
XX
AC   AAC40589;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 28855.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
```

PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

```
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match           95.5%;  Score 1002.2;  DB 21;  Length 1129;
  Best Local Similarity   97.8%;  Pred. No. 1.5e-283;
  Matches 1016;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy         11 TAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATCGACGACGC 70
              | | |||  ||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          2 TTAATTTGTTTTTTTTGTTTCATTAAAAGAAATGTCATCGTCAAATTGGATCGACGACGC 61

Qy         71 TTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCT 121

Qy        131 TTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGG 181

Qy        191 CCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCAT 250
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        182 CCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTACCTCGTTCCAT 241

Qy        251 CACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATTTTCC 310
              ||||| | |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        242 CACTTGTCCTACATCTTCTAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATTTTCC 301

Qy        311 AGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCG 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 AGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCG 361

Qy        371 AGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGT 430
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        362 AGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAAGAGCTTTTGT 421

Qy        431 TGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGC 490
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TGGCTTTGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGC 481

Qy        491 GACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGG 550
              ||||||||||||||||||||  ||||||||||||||||||||||||| ||||||||
Db        482 AACTGTCCAGATATGTGTAGATAATAATTATTGTGATGTTATGCATATTTTCCATTCTGG 541

Qy        551 TATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGG 610
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        542 TATCCCTCAAAGTCTCCAACATCTTATTGAGGATTCAACACTTGTAAAGGTAGGTATTGG 601

Qy        611 AATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 AATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGA 661

Qy        671 GGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTC 730
              |||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||
Db        662 GGATCTTTCAGATTTAGCCAACCAAAAAATTGATGGAGAGAAAAAATGGGGCCTTGCCTC 721

Qy        731 ACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAA 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 ACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAA 781

Qy        791 CTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGC 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 CTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGC 841

Qy        851 TTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTGGCTCATAACG 910
              |||||||||||||||||||||||||||||||||||||||||||||||| | |||||||||
Db        842 TTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGCGACTCATAACG 901

Qy        911 TGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAAATGATATGAT 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 TGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAAATGATATGAT 961

Qy        971 ACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGTTCTGGTCCCT 1030
              |||||||||| |||||||||||||||||||||||||||| ||||||||||||||||||
Db        962 ACACCTAATCCAGTCAAGTAGATGCAATTCTTGTGAATATTTTATCTAGTTCTGGTCCCT 1021

Qy       1031 TTAACCGTCCAGAAACTAG 1049
              |||| ||||||||||||||
Db       1022 TTAATCGTCCAGAAACTAG 1040


RESULT 4
ABA91795
ID   ABA91795 standard; cDNA; 942 BP.
XX
AC   ABA91795;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Arabidopsis RNase D related domain (3'-5' exonuclease domain) DNA.
XX
KW   RNase D; enzyme; 3'-5' exonuclease; gene expression; plant;
KW   gene silencing; transgenic plant; gene; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..942
FT                   /*tag= a
FT                   /product= "3'-5'_exonuclease_domain"
XX
PN   WO200210362-A2.
XX
PD   07-FEB-2002.
XX
PF   30-JUL-2001; 2001WO-EP08825.
XX
PR   01-AUG-2000; 2000US-222202P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA   (NOVS ) NOVARTIS FORSCHUNGSSTIFTUNG ZWEIGNIEDERL.
XX
PI   Levin JZ,  Phillips KL,  Budziszewski GJ,  Meins F,  Glazov EA;
XX
DR   WPI; 2002-217119/27.
DR   P-PSDB; AAM50927.
XX
PT   Regulating gene expression in plants for controlling gene silencing,
PT   comprises altering the transcription or translation of an endonuclease
PT   nucleotide sequence encoding a polypeptide comprising an exonuclease
PT   domain -
XX
PS   Claim 4; Page 70; 102pp; English.
XX
CC   The present nucleotide sequence encodes an Arabidopsis thaliana 3'-5'
CC   exonuclease domain (see AAM50927), specifically an RNase D related
CC   domain (GenPept accession CAB36851).  The nucleotide sequence is
CC   found in BAC F18A5, GenBank accession number AL035528.2.  Nucleotide
CC   (see ABA91795-807) and polypeptide (see AAM50927-39) sequences for
CC   3'-5' exonuclease domains were identified using different screening
CC   methods and search algorithms.  The invention encompasses the
CC   suppression or increase of gene silencing in plants.  This is
CC   achieved by altering the expression in the plant cell of a
CC   nucleotide sequence encoding a polypeptide comprising an exonuclease
CC   domain, preferably a 3'-5' exonuclease domain, and especially an
CC   RNase D-related domain.  Expression of the nucleotide sequence is
```

```
CC   altered by altering its transcription or translation.  It is
CC   reduced e.g. by sense suppression, antisense suppression, homologous
CC   recombination, the use of ribozymes, dominant-negative mutants,
CC   aptamers, zinc finger proteins, double-stranded RNA, or insertional,
CC   point or deletion mutagenesis, and increased by overexpression.
CC   The methods are useful for altering or stabilizing the expression of
CC   a nucleotide sequence of interest in a plant cell or plant.
CC   Suppressing or decreasing expression of the nucleic acid molecule
CC   results in decreased levels of post transcriptional gene silencing
CC   and improved expression of genes of interest.
XX
SQ   Sequence 942 BP; 263 A; 203 C; 195 G; 281 T; 0 other;

  Query Match             79.2%;  Score 831;  DB 24;  Length 942;
  Best Local Similarity   96.1%;  Pred. No. 2.4e-233;
  Matches  852;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         42 ATGTCATCGTCAAATTGGATCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGAC 101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCATCGTCAAATTGGATCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGAC 60

Qy        102 GCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCG 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCG 120

Qy        162 ACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAAT 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAAT 180

Qy        222 AATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCT 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCT 240

Qy        282 CTCTCCCGTTGCCGAGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTAT 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCTCCCGTTGCCGAGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTAT 300

Qy        342 AGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACC 401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACC 360

Qy        402 AAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTT 461
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTT 420

Qy        462 AGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTAT 521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTAT 480

Qy        522 TGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAA 581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAA 540

Qy        582 GATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCAT 641
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCAT 600

Qy        642 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 701
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 660

Qy        702 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 720

Qy        762 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 780

Qy        822 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGAC 881
              ||||||||||||||||||||||||||||||||||||||||||||||||||    |
Db        781 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTAACAACAACG 840

Qy        882 CTTCCTGATGCTGTCAGTGGCTCATAACGTGAAGGAGGAAGCTTAAA 928
                  ||  | |||    |    | | |||||  |  || || | |
Db        841 AAAAACCATCTTCTCACACTCAACGACCTTGAAGCAAAAATCTCACA 887


RESULT 5
ABA91803
ID   ABA91803 standard; cDNA; 4299 BP.
XX
AC   ABA91803;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Human Werner syndrome protein (3'-5' exonuclease domain) DNA.
XX
KW   Enzyme; 3'-5' exonuclease; human; gene expression;
KW   Werner syndrome protein; gene silencing; transgenic plant; gene;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..4299
FT                   /*tag= a
FT                   /product= "3'-5'_exonuclease_domain"
XX
PN   WO200210362-A2.
XX
PD   07-FEB-2002.
XX
PF   30-JUL-2001; 2001WO-EP08825.
XX
PR   01-AUG-2000; 2000US-222202P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA   (NOVS ) NOVARTIS FORSCHUNGSSTIFTUNG ZWEIGNIEDERL.
XX
PI   Levin JZ,  Phillips KL,  Budziszewski GJ,  Meins F,  Glazov EA;
XX
DR   WPI; 2002-217119/27.
DR   P-PSDB; AAM50935.
XX
PT   Regulating gene expression in plants for controlling gene silencing,
PT   comprises altering the transcription or translation of an endonuclease
PT   nucleotide sequence encoding a polypeptide comprising an exonuclease
PT   domain  -
XX
PS   Claim 42; Page 85-86; 102pp; English.
XX
CC   The present nucleotide sequence is that of human DNA encoding
CC   Werner syndrome protein (see AAM50935), the N-terminal portion of
CC   which comprises a 3'-5' exonuclease domain.  The sequence was
CC   identified on the basis of homology to an Arabidopsis 3'-5'
CC   exonuclease domain (see AAM50927).  The invention provides nucleotide
CC   (see ABA91795-807) and polypeptide (see AAM50927-39) sequences for
CC   3'-5' exonuclease domains identified using different screening
CC   methods and search algorithms.  The invention encompasses the
CC   suppression or increase of gene silencing in plants.  This is
CC   achieved by altering the expression in the plant cell of a
CC   nucleotide sequence encoding a polypeptide comprising an exonuclease
CC   domain, preferably a 3'-5' exonuclease domain, and especially an
CC   RNase D-related domain.  Expression of the nucleotide sequence is
CC   modulated by altering its transcription or translation.  It is
CC   reduced e.g. by sense suppression, antisense suppression, homologous
CC   recombination, the use of ribozymes, dominant-negative mutants,
CC   aptamers, zinc finger proteins, double-stranded RNA, or insertional,
CC   point or deletion mutagenesis, and increased by overexpression.
CC   The methods are useful for altering or stabilizing the expression of
CC   a nucleotide sequence of interest in a plant cell or plant.
CC   Suppressing or decreasing expression of the nucleic acid molecule
CC   results in decreased levels of post transcriptional gene silencing
CC   and improved expression of genes of interest.
XX
```

```
SQ   Sequence 4299 BP; 1400 A; 759 C; 950 G; 1190 T; 0 other;

  Query Match            10.6%;  Score 111;  DB 24;  Length 4299;
  Best Local Similarity  54.1%;  Pred. No. 9.1e-22;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || ||||||  ||||
Db        202 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 261

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||    || || |||||     | ||| | |||||   |
Db        262 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 318

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        319 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 378

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
Db        379 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 438

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  ||||| | |||||
Db        439 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 498

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |     |  | | ||| |||||  |   ||   | |    ||  | |  |
Db        499 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 558

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||   ||    ||| |   | | || |||      | ||||||  |
Db        559 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 618

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        619 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 678

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        679 TTAGAGATTTTGGATGATACTGT 701


RESULT 6
AAV65701
ID   AAV65701 standard; cDNA to mRNA; 5189 BP.
XX
AC   AAV65701;
XX
DT   10-DEC-1998  (first entry)
XX
DE   DNA fragment of a Werner's syndrome gene.
XX
KW   Werner's syndrome; diagnosis; ds.
XX
OS   Homo sapiens.
XX
PN   JP10201498-A.
XX
PD   04-AUG-1998.
XX
PF   24-JAN-1997;   97JP-0011268.
XX
PR   24-JAN-1997;   97JP-0011268.
XX
PA   (EIJI-) EIJIN KENKYUSHO KK.
XX
DR   WPI; 1998-474499/41.
XX
PT   Detection of mutation in gene causing human Werner's syndrome - and
PT   oligo:nucleotide used for detection, comprises amplifying DNA and
PT   synthesising oligo:nucleotide
XX
PS   Claim 1; Pages 10-11; 17pp; Japanese.
XX
CC   The present sequence represents a gene, the mutation of which causes
CC   Werner's syndrome. The specification describes the detection of a
CC   mutation in a gene causing human Werner's syndrome. The method comprises
CC   amplifying a DNA fragment containing a mutation at position 733, 734,
CC   1620 or 4146 of AAV65701 or at position 42 of AAV65702 and synthesising
CC   an oligonucleotide so that the base at the above site comes to be the 3'
CC   end based on the base sequence of AAV65701-02, or an oligonucleotide in
CC   which the base adjacent to the 3' end comes to be the 5' end. The
CC   oligonucleotides are hybridised with the resultant amplified fragment.
CC   The method can be used to diagnose Werner's syndrome.
XX
SQ   Sequence 5189 BP; 1663 A; 889 C; 1129 G; 1508 T; 0 other;

  Query Match            10.6%;  Score 111;  DB 19;  Length 5189;
  Best Local Similarity  54.1%;  Pred. No. 9.9e-22;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || ||||||  ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||    || || |||||     | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  ||||| | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |     |  | | ||| |||||  |   ||   | |    ||  | |  |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||   ||    ||| |   | | || |||      | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932


RESULT 7
ABL65821
ID   ABL65821 standard; DNA; 5189 BP.
XX
AC   ABL65821;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Lung cancer related gene sequence SEQ ID NO:4158.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
```

PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 4158; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 5189 BP; 1663 A; 889 C; 1129 G; 1508 T; 0 other;

Query Match 10.6%; Score 111; DB 24; Length 5189;
Best Local Similarity 54.1%; Pred. No. 9.9e-22;
Matches 272; Conservative 0; Mismatches 225; Indels 6; Gaps 2;

```
Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || |||||| ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  | | |||    ||    || || |||||     | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |       || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||     |||
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  ||||| | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |     |     | |  ||| |||||  |   ||   | |    ||  | |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||    ||    ||| |   | | || |||      | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932
```

RESULT 8
AAX83001
ID AAX83001 standard; DNA; 5208 BP.
XX
AC AAX83001;
XX
DT 31-AUG-1999 (first entry)
XX
DE Human WRN coding sequence.
XX
KW Human; WRN; Werner's syndrome; detection; diagnosis; autosomal;
KW recessive disorder; phenotype; ss.
XX
OS Homo sapiens.
XX
PN WO9724435-A1.
XX
PD 10-JUL-1997.
XX
PF 30-DEC-1996; 96WO-US20785.
XX
PR 12-APR-1996; 96US-0632175.
PR 29-DEC-1995; 95US-0009409.

```
PR   29-DEC-1995;   95US-0580539.
PR   30-JAN-1996;   96US-0010835.
PR   30-JAN-1996;   96US-0594242.
XX
PA   (DARW-) DARWIN MOLECULAR CORP.
PA   (OSHI/) OSHIMA J.
XX
PI   Fu Y,  Mulligan J,  Oshima J,  Schellenberg GD,  Yu C;
XX
DR   WPI; 1997-363671/33.
DR   P-PSDB; AAY14517.
XX
PT   Isolated nucleic acid molecule encoding the WRN gene product -
PT   useful for detection and treatment of Werner's syndrome, and related
PT   diseases
XX
PS   Claim 1; Fig 2A; 153pp; English.
XX
CC   This sequence represents the coding region for the human WRN gene which
CC   encodes a protein related to Werner's syndrome.  The products can be used
CC   for the detection and treatment of Werner's syndrome (WS), an autosomal
CC   recessive disorder with a complex phenotype, as well as related diseases.
XX
SQ   Sequence 5208 BP; 1682 A; 889 C; 1129 G; 1508 T; 0 other;

  Query Match             10.6%;  Score 111;  DB 18;  Length 5208;
  Best Local Similarity   54.1%;  Pred. No. 9.9e-22;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || |||||  ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||    || || |||||     | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              | |||||  ||  | || ||  ||||| |||| | |||||||| || ||      |||
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  |||||  | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |     |     | |  ||| |||||  |   ||   | |    ||  |  |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||    ||    ||| |   | | || |||      | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
               || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932


RESULT 9
AAX24302
ID   AAX24302 standard; DNA; 6476 BP.
XX
AC   AAX24302;
XX
DT   07-JUN-1999  (first entry)
XX
DE   Murine mWRN gene (WRN homologue).
XX
KW   mWRN gene; WRN gene; homologue; Werner's syndrome; mouse;
KW   life span; ageing; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     229..4434
FT                   /*tag= a
FT   polyA_signal    5812..5817
FT                   /*tag= b
FT   polyA_signal    5917..5922
FT                   /*tag= c
FT   polyA_signal    6411..6416
FT                   /*tag= d
XX
PN   WO9906543-A1.
XX
PD   11-FEB-1999.
XX
PF   03-AUG-1998;   98WO-US16081.
XX
PR   04-AUG-1997;   97US-0054629.
XX
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI   Guarente LP,  Lombard DB,  Sinclair DA;
XX
DR   WPI; 1999-153773/13.
DR   P-PSDB; AAW97841.
XX
PT   Identifying agents that inhibit replication and accumulation of DNA
PT   circles - as potential agents for increasing the life span of
PT   animals
XX
PS   Claim 17; Page 62-69; 78pp; English.
XX
CC   This is the nucleotide sequence of the mWRN gene, the murine
CC   homologue of the human WRN gene associated with Werner's syndrome.
CC   A claimed assay for identifying a compound which extends life span
CC   comprises administering a compound to be tested to a mammal with a
CC   suppressed level of mWRN, and identifying slowing of at least one
CC   of the normal phenotypes of ageing in the mammal.  The compound
CC   identified by this method inhibits the replication and/or
CC   accumulation of rDNA circles in a cell.  The assay can be
CC   performed using a mWRN knockout mouse (also claimed).  mWRN nucleic
CC   acids can be used to produce mWRN polypeptides and as diagnostic
CC   probes and primers.  mWRN polypeptides can be used to raise
CC   antibodies (for inhibiting activity of mWRN or for determining it
CC   in tissues).  The assay is based on the observation that in yeast
CC   cells, accumulation of ribosomal circular DNA is responsible for
CC   age-related enlargement and fragmentation of the nucleolus.  A
CC   mutation in the yeast WRN homologue SGS1 causes premature ageing,
CC   suggesting a common ageing mechanism in all eukaryotes.  Insight
CC   into the ageing process in model systems can provide insight into
CC   ageing in humans.
XX
SQ   Sequence 6476 BP; 2019 A; 1300 C; 1389 G; 1768 T; 0 other;

  Query Match             7.9%;  Score 82.8;  DB 20;  Length 6476;
  Best Local Similarity   53.2%;  Pred. No. 2.1e-13;
  Matches  199;  Conservative    0;  Mismatches  172;  Indels    3;  Gaps    1;

Qy        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540
              | |  || ||     |||||  | |||||   |   ||  | |||  |  |  | || ||||
Db        494 GCAGAGTCGCAGTGATCCAGTTGTGTGTGTCTGAGAACAAATGTTACTTGTTTCACATTT 553

Qy        541 TT---CATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAA 597
                  |      || | | |||| ||  |  |  ||    |  | |||  |   | || ||   |
Db        554 CTTCCATGTCAGTTTTCCCCCAGGGATTAAAAATGTTACTAGAAAACAAATCAATTAAGA 613
```

```
Qy        598 AGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTA 657
              ||| |||  |||| ||||| || |||     ||| ||| | | ||| | ||  || |
Db        614 AGGCAGGGGTTGGGATTGAAGGGACCAGTGGAAACTTCTGCGTGATTTTGACGTCAAGT 673

Qy        658 TCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAAT 717
              |  |   | ||| ||| ||  | ||| | ||||  ||||  |     | |   | |  |
Db        674 TGGAGAGTTTTGTGGAGCTGACGGATGTTGCCAATGAAAAGTTGAAGTGCGCAGAGACCT 733

Qy        718 GGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAA 777
              || ||||       ||   | |     |  |  | ||| | ||  ||||    ||    |
Db        734 GGAGCCTCAATGGTCTGGTTAAACACGTCTTAGGGAAACAACTTTTGAAAGACAAGTCCA 793

Qy        778 TCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAA 837
              || |     | || |||     ||  || ||  |  || | |||  ||  || ||||| |
Db        794 TCCGCTGCAGCAATTGGAGTAATTTCCCCCTCACTGAGGACCAGAAACTGTATGCAGCCA 853

Qy        838 CGGATGCTTATGCT 851
              | ||||||||||||
Db        854 CTGATGCTTATGCT 867
```

RESULT 10
AAV35131
ID AAV35131 standard; cDNA to mRNA; 1164 BP.
XX
AC AAV35131;
XX
DT 28-AUG-1998 (first entry)
XX
DE Mouse WRN helicase RACE fragment cDNA #3.
XX
KW Werner's syndrome; WRN; helicase; murine; diagnosis; RACE; ds.
XX
OS Mus sp.
XX
PN JP10146188-A.
XX
PD 02-JUN-1998.
XX
PF 15-NOV-1996; 96JP-0304721.
XX
PR 15-NOV-1996; 96JP-0304721.
XX
PA (EIJI-) EIJIN KENKYUSHO KK.
XX
DR WPI; 1998-369863/32.
XX
PT New mouse gene encoding protein - used for diagnosis of human
PT Werner's syndrome
XX
PS Example 1; Page 26-27; 30pp; Japanese.
XX
CC This sequence is a fragment of novel mouse WRN helicase protein which
CC is associated with Werner's syndrome and has been generated by RACE
CC (rapid amplification of cDNA ends). The mouse WRN gene is useful for
CC the diagnosis of human Werner's syndrome.
XX
SQ Sequence 1164 BP; 343 A; 223 C; 288 G; 310 T; 0 other;

Query Match 7.7%; Score 81.2; DB 19; Length 1164;
Best Local Similarity 52.9%; Pred. No. 2.7e-13;
Matches 198; Conservative 0; Mismatches 173; Indels 3; Gaps 1;

```
Qy        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540
              | |  || ||     ||||| | |||||    |   ||  | ||| |  |  | || ||||
Db        449 GCAGAGTCGCAGTGATCCAGTTGTGTGTGTCTGAGAACAAATGTTACTTGTTTCACATTT 508

Qy        541 TT---CATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAA 597
                  ||  | | |||| ||  |  |  ||    |  | ||| |   | || ||   |
Db        509 CTTCCATGTCAGTTTTCCCCCAGGGATTAAAAATGTTACTAGAAAACAAATCAATTAAGA 568

Qy        598 AGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTA 657
              ||| |||  |||| ||||| || |||     ||| ||| | | ||| | ||  || |
Db        569 AGGCAGGGGTTGGGATTGAAGGGACCAGTGGAAACTTCTGCGTGATTTTGACGTCAAGT 628

Qy        658 TCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAAT 717
              |  |   | ||| ||| ||  | ||| | |||||  ||||  |     | |   | |  |
Db        629 TGGAGAGTTTTGTGGAGCTGACGGATGTTGCCAATGAAAAGTTGAAGTGCGCAGAGACCT 688

Qy        718 GGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAA 777
              || ||||       ||   | |     |  |  | ||| | ||  ||||    ||    |
Db        689 GGAGCCTCAATGGTCTGGTTAAACACGTCTTAGGGAAACAACTTTTGAAAGACAAGTCCA 748

Qy        778 TCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAA 837
              || |     | || |||     ||  || ||  |  || | |||  ||  || ||||| |
Db        749 TCCGCTGCAGCAATTGGAGTAATTTCCCCCTCACTGAGGACCAGAAACTGTATGCAGCCA 808

Qy        838 CGGATGCTTATGCT 851
              | ||||| ||||||
Db        809 CTGATGCCTATGCT 822
```

RESULT 11
AAV35115
ID AAV35115 standard; cDNA to mRNA; 4206 BP.
XX
AC AAV35115;
XX
DT 28-AUG-1998 (first entry)
XX
DE Mouse WRN helicase cDNA.
XX
KW Werner's syndrome; WRN; helicase; murine; diagnosis; ds.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..4206
FT /*tag= a
FT /product= WRN helicase
XX
PN JP10146188-A.
XX
PD 02-JUN-1998.
XX
PF 15-NOV-1996; 96JP-0304721.
XX
PR 15-NOV-1996; 96JP-0304721.
XX
PA (EIJI-) EIJIN KENKYUSHO KK.
XX
DR WPI; 1998-369863/32.
DR P-PSDB; AAW59454.
XX
PT New mouse gene encoding protein - used for diagnosis of human
PT Werner's syndrome
XX
PS Claim 2; Page 22-23; 30pp; Japanese.
XX
CC This sequence encodes a novel mouse WRN helicase protein which is
CC associated with Werner's syndrome. The mouse WRN gene is useful for
CC the diagnosis of human Werner's syndrome.
XX
SQ Sequence 4206 BP; 1284 A; 876 C; 976 G; 1070 T; 0 other;

Query Match 7.7%; Score 81.2; DB 19; Length 4206;
Best Local Similarity 52.9%; Pred. No. 5e-13;
Matches 198; Conservative 0; Mismatches 173; Indels 3; Gaps 1;

```
Qy        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540
              | |  || ||     ||||| | |||||    |   ||  | ||| |  |  | || ||||
Db        266 GCAGAGTCGCAGTGATCCAGTTGTGTGTGTCTGAGAACAAATGTTACTTGTTTCACATTT 325

Qy        541 TT---CATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAA 597
                  ||  | | |||| ||  |  |  ||    |  | ||| |   | || ||   |
Db        326 CTTCCATGTCAGTTTTCCCCCAGGGATTAAAAATGTTACTAGAAAACAAATCAATTAAGA 385
```

```
Qy        598 AGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTA 657
              ||| |||  |||| |||||| || |||     ||| ||| | | ||| | ||  || |
Db        386 AGGCAGGGGTTGGGATTGAAGGGGACCAGTGGAAACTTCTGCGTGATTTTGACGTCAAGT 445

Qy        658 TCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAAT 717
              |  |   | ||| ||| ||  | ||| | |||||  |||  |     | |   | |  |
Db        446 TGGAGAGTTTTGTGGAGCTGACGGATGTTGCCAATGAAAAGTTGAAGTGCGCAGAGACCT 505

Qy        718 GGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAA 777
              || ||||       ||   | |      |  |  | ||| | ||  ||||    ||    |
Db        506 GGAGCCTCAATGGTCTGGTTAAACACGTCTTAGGGAAACAACTTTTGAAAGACAAGTCCA 565

Qy        778 TCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAA 837
              || |     | || |||     ||  || ||  |  || | |||  ||  || ||||| |
Db        566 TCCGCTGCAGCAATTGGAGTAATTTCCCCCTCACTGAGGACCAGAAACTGTATGCAGCCA 625

Qy        838 CGGATGCTTATGCT 851
              | ||||| ||||||
Db        626 CTGATGCCTATGCT 639
```

RESULT 12
AAX83004
ID AAX83004 standard; DNA; 4792 BP.
XX
AC AAX83004;
XX
DT 31-AUG-1999 (first entry)
XX
DE Mouse WRN coding sequence.
XX
KW Mouse; WRN; Werner's syndrome; detection; diagnosis; autosomal;
KW recessive disorder; phenotype; ss.
XX
OS Mus musculus.
XX
PN WO9724435-A1.
XX
PD 10-JUL-1997.
XX
PF 30-DEC-1996; 96WO-US20785.
XX
PR 12-APR-1996; 96US-0632175.
PR 29-DEC-1995; 95US-0009409.
PR 29-DEC-1995; 95US-0580539.
PR 30-JAN-1996; 96US-0010835.
PR 30-JAN-1996; 96US-0594242.
XX
PA (DARW-) DARWIN MOLECULAR CORP.
PA (OSHI/) OSHIMA J.
XX
PI Fu Y, Mulligan J, Oshima J, Schellenberg GD, Yu C;
XX
DR WPI; 1997-363671/33.
DR P-PSDB; AAY14519.
XX
PT Isolated nucleic acid molecule encoding the WRN gene product -
PT useful for detection and treatment of Werner's syndrome, and related
PT diseases
XX
PS Claim 1; Fig 6; 153pp; English.
XX
CC This sequence represents the coding region for the mouse WRN gene. The
CC corresponding human gene (AAX83001) encodes a protein related to
CC Werner's syndrome. The products can be used for the detection and
CC treatment of Werner's syndrome (WS), an autosomal recessive disorder with
CC a complex phenotype, as well as related diseases.
XX
SQ Sequence 4792 BP; 1441 A; 1012 C; 1107 G; 1226 T; 6 other;

Query Match 7.7%; Score 81.2; DB 18; Length 4792;
Best Local Similarity 52.3%; Pred. No. 5.3e-13;
Matches 179; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

```
Qy        510 GATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAA 569
              || || || | ||       ||     |  ||     || | | |||| ||  |  |  ||
Db        442 GAGAGCAAATGTTACTTGTTTCACATTTCTTCCATGTCAGTTTTCCCCCAGGGATTAAAA 501

Qy        570 CATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTG 629
                  |  | ||| |   | || ||   |||| |||  |||| ||||| || |||     |
Db        502 ATGTTACTAGAAAACAAATCAATTAAGAAGGCAGGGGTTGGGATTGAAGGGGACCAGTGG 561

Qy        630 AAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCC 689
              || ||| | | ||| | ||  || |   |  |   | ||| ||| ||  | ||| | |||
Db        562 AAACTTCTGCGTGATTTTGACGTCAAGTTGGAGAGTTTTGTGGAGCTGACGGATGTTGCC 621

Qy        690 AACCAAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTT 749
              ||  |||| |     | |   | |  ||| ||||         ||   | |     |  |
Db        622 AATGAAAAGTTGAAGTGCGCAGAGACCTGGAGCCTCAATGGTCTGGTTAAACACGTCTTA 681

Qy        750 TGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTG 809
              | ||| | ||  ||||    ||    ||| |     | || |||     ||  || ||
Db        682 GGGAAACAACTTTTGAAAGACAAGTCCATCCGCTGCAGCAATTGGAGTAATTTCCCCCTC 741

Qy        810 TCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCT 851
              |  || | |||  ||  || ||||| || ||||||||||||
Db        742 ACTGAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCT 783
```

RESULT 13
AAV35114
ID AAV35114 standard; cDNA to mRNA; 5058 BP.
XX
AC AAV35114;
XX
DT 28-AUG-1998 (first entry)
XX
DE Mouse WRN helicase cDNA including non-coding region.
XX
KW Werner's syndrome; WRN; helicase; murine; diagnosis; ds.
XX
OS Mus sp.
XX
FH Key   Location/Qualifiers
FT CDS   184..4389
FT   /*tag= a
FT   /product= WRN helicase
XX
PN JP10146188-A.
XX
PD 02-JUN-1998.
XX
PF 15-NOV-1996; 96JP-0304721.
XX
PR 15-NOV-1996; 96JP-0304721.
XX
PA (EIJI-) EIJIN KENKYUSHO KK.
XX
DR WPI; 1998-369863/32.
DR P-PSDB; AAW59454.
XX
PT New mouse gene encoding protein - used for diagnosis of human
PT Werner's syndrome
XX
PS Claim 3; Page 13-18; 30pp; Japanese.
XX
CC This sequence encodes a novel mouse WRN helicase protein which is
CC associated with Werner's syndrome. The mouse WRN gene is useful for
CC the diagnosis of human Werner's syndrome.
XX
SQ Sequence 5058 BP; 1546 A; 1034 C; 1156 G; 1322 T; 0 other;

Query Match 7.7%; Score 81.2; DB 19; Length 5058;
Best Local Similarity 52.9%; Pred. No. 5.4e-13;
Matches 198; Conservative 0; Mismatches 173; Indels 3; Gaps 1;

```
Qy        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540
              | |  || ||     |||||  | |||||    |   ||  | ||| |  |  | || ||||
Db        449 GCAGAGTCGCAGTGATCCAGTTGTGTGTGTCTGAGAACAAATGTTACTTGTTTCACATTT 508

Qy        541 TT---CATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAA 597
               |      || | | |||| ||  |  |  ||    |  | ||| |   | || ||   |
Db        509 CTTCCATGTCAGTTTTCCCCCAGGGATTAAAAATGTTACTAGAAAACAAATCAATTAAGA 568

Qy        598 AGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTA 657
              ||| |||  |||| ||||| || |||     ||| ||| | | ||| | ||  || |
Db        569 AGGCAGGGGTTGGGATTGAAGGGGACCAGTGGAAACTTCTGCGTGATTTTGACGTCAAGT 628

Qy        658 TCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAAT 717
              |  |   | ||| ||| ||  | ||| | |||||  ||||  |     | |   | |  |
Db        629 TGGAGAGTTTTGTGGAGCTGACGGATGTTGCCAATGAAAAGTTGAAGTGCGCAGAGACCT 688

Qy        718 GGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAA 777
              || ||||       ||   | |     |  |  | ||| | ||  ||||    ||    |
Db        689 GGAGCCTCAATGGTCTGGTTAAACACGTCTTAGGGAAACAACTTTTGAAAGACAAGTCCA 748

Qy        778 TCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAA 837
              || |     | || |||     ||  || ||  |  || | |||  ||  || ||||| |
Db        749 TCCGCTGCAGCAATTGGAGTAATTTCCCCCTCACTGAGGACCAGAAACTGTATGCAGCCA 808

Qy        838 CGGATGCTTATGCT 851
              | ||||| ||||||
Db        809 CTGATGCCTATGCT 822
```

RESULT 14
AAX83005
ID AAX83005 standard; DNA; 29604 BP.
XX
AC AAX83005;
XX
DT 31-AUG-1999 (first entry)
XX
DE Partial mouse WRN genomic sequence #1.
XX
KW Mouse; WRN; Werner's syndrome; detection; diagnosis; autosomal;
KW recessive disorder; phenotype; ss.
XX
OS Mus musculus.
XX
PN WO9724435-A1.
XX
PD 10-JUL-1997.
XX
PF 30-DEC-1996; 96WO-US20785.
XX
PR 12-APR-1996; 96US-0632175.
PR 29-DEC-1995; 95US-0009409.
PR 29-DEC-1995; 95US-0580539.
PR 30-JAN-1996; 96US-0010835.
PR 30-JAN-1996; 96US-0594242.
XX
PA (DARW-) DARWIN MOLECULAR CORP.
PA (OSHI/) OSHIMA J.
XX
PI Fu Y, Mulligan J, Oshima J, Schellenberg GD, Yu C;
XX
DR WPI; 1997-363671/33.
XX
PT Isolated nucleic acid molecule encoding the WRN gene product -
PT useful for detection and treatment of Werner's syndrome, and related
PT diseases
XX
PS Claim 1; Fig 7; 153pp; English.
XX
CC This sequence represents a fragment of the genomic sequence containing
CC the coding region for the mouse WRN gene (AAX83004). The corresponding
CC human gene (AAX83001) encodes a protein related to Werner's syndrome.
CC The products can be used for the detection and treatment of Werner's
CC syndrome (WS), an autosomal recessive disorder with a complex phenotype,
CC as well as related diseases.
XX
SQ Sequence 29604 BP; 7634 A; 5861 C; 5985 G; 10123 T; 1 other;

Query Match 4.9%; Score 51.2; DB 18; Length 29604;
Best Local Similarity 56.5%; Pred. No. 0.0008;
Matches 95; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy          545 TTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGG 604
                ||  | | |||| ||  |  |  ||    |  | ||| |   | || ||   |||| |||
Db        19789 TTTAGTTTTCCCCCAGGGATTAAAAATGTTACTAGAAAACAAATCAATTAAGAAGGCAGG 19848

Qy          605 TATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGA 664
                  |||| ||||| || |||     ||| ||| | | ||| | ||  || |   |  |
Db        19849 GGTTGGGATTGAAGGGGACCAGTGGAAACTTCTGCGTGATTTTGACGTCAAGTTGGAGAG 19908

Qy          665 TGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAA 712
                | ||| ||| ||  | ||| | ||||  ||||  | || | |   ||
Db        19909 TTTTGTGGAGCTGACGGATGTTGCCAATGAAAAGGTAGGCGTAATAAA 19956
```

RESULT 15
ABA08508
ID ABA08508 standard; cDNA; 449 BP.
XX
AC ABA08508;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human secreted protein homologue-encoding cDNA, SEQ ID NO:284.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnerary; antiulcer; ss.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US03800.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-457740/49.
DR P-PSDB; ABB11264.
XX
PT Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
PS Claim 1; Page 436; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The

CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention.
XX
SQ Sequence 449 BP; 113 A; 101 C; 112 G; 118 T; 5 other;

Query Match 4.2%; Score 44.4; DB 22; Length 449;
Best Local Similarity 49.3%; Pred. No. 0.011;
Matches 148; Conservative 0; Mismatches 146; Indels 6; Gaps 1;

```
Qy        571 ATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGA 630
              || || | | ||||   ||  || | || || ||  | |||      |  ||  |    |
Db         27 ATATTTTGGCAGATGGCACCATTTTGAAAGTTGGAGTGGGATGCTCAGAAGATGCCAGCA 86

Qy        631 AGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCA 690
              ||||| | || || |||||  |   | | |  |       ||| ||   | |  ||||||
Db         87 AGCTTCTGCAGGATTATGGCCTCGTTGTTAGGGGGTGCCTGGACCTCCGATACCTAGCCA 146

Qy        691 ------ACCAAAAAAATTGGTGGAGATAAAAAAATGGGGCCTTGCCTCACTAACTGAGACAC 744
                  |   ||| | |         |||         ||||    || ||  |||||||
Db        147 TGCGGCAGAGAAACAATTTGCTCTGTAATGGGCTTAGCCTGAAGTCCCTCGCTGAGACTG 206

Qy        745 TTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATC 804
              || |   |       ||    ||| |      | |  |     | ||||||||   | |
Db        207 TTTTGAACTTTCCCCTTGACAAGTCCCTTCTACTTCGTTGCAGCAACTGGGATGCTGAGA 266

Qy        805 CTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTT 864
              ||||  || || | ||| ||   || || || | ||||||  |   ||||  |  ||| |
Db        267 CTCTCACAGAGGACCAGGTAATTTATGCTGCCAGGGATGCCCAGATTTCAGTGGCTCTCT 326
```

Search completed: June 14, 2003, 07:03:19
Job time : 322 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: June 14, 2003, 06:36:38 ; Search time 1820 Seconds
(without alignments)
9334.651 Million cell updates/sec

Title: US-09-896-186B-23
Perfect score: 1049
Sequence: 1 accaaagcattaatttttat.....tttaaccgtccagaaactag 1049

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	288.8	27.5	435	9	AU226180	AU226180 AU226180
c	2	273.8	26.1	840	10	BE659005	BE659005 GM700008A
	3	261	24.9	704	12	BG441595	BG441595 GA__Ea001
	4	245	23.4	552	10	AW201789	AW201789 sf07c10.y
	5	243.8	23.2	671	9	AU237507	AU237507 AU237507
	6	241.2	23.0	467	10	AW832139	AW832139 sm20b11.y
	7	235.4	22.4	661	14	BQ799068	BQ799068 EST 1237
	8	233.8	22.3	535	13	BM406124	BM406124 EST580451
	9	223	21.3	506	12	BG133220	BG133220 EST466112
	10	213.6	20.4	531	13	BM094501	BM094501 saj16g03.
	11	172.2	16.4	515	13	BM358892	BM358892 GA__Ea001
	12	169.6	16.2	517	13	BM358891	BM358891 GA__Ea001
	13	167.4	16.0	555	14	BM731125	BM731125 sal68a05.
	14	162	15.4	595	13	BJ213300	BJ213300 BJ213300
	15	156.8	14.9	499	13	BI433038	BI433038 EST535799
	16	156.6	14.9	556	14	BM890816	BM890816 sam08c02.
	17	156	14.9	563	13	BM527093	BM527093 sal49e10.
	18	152	14.5	531	10	AW460204	AW460204 sh56c10.y
	19	151	14.4	684	12	BG589996	BG589996 EST497838
	20	143.2	13.7	625	12	BE999775	BE999775 EST431498
	21	142.4	13.6	560	12	BG157629	BG157629 saa88c03.
	22	140.8	13.4	572	9	AU235472	AU235472 AU235472
c	23	134.4	12.8	680	13	BJ321563	BJ321563 BJ321563
c	24	129.6	12.4	407	12	BE821184	BE821184 GM700024A
c	25	129	12.3	220	17	BH849559	BH849559 SALK_0698
	26	124.6	11.9	431	9	AI748378	AI748378 sb52a06.y
	27	117	11.2	445	10	AW119478	AW119478 sd46a06.y
	28	100.6	9.6	494	10	AW720170	AW720170 LjNEST16b
c	29	99	9.4	654	13	BJ220795	BJ220795 BJ220795
c	30	97.8	9.3	660	17	B61114	B61114 T20K8TF TAM
	31	96	9.2	600	13	BJ316072	BJ316072 BJ316072
	32	84.6	8.1	550	14	BM785803	BM785803 K-EST0064
	33	76.2	7.3	405	17	CNS00V8H	AL091967 Arabidops
	34	66.8	6.4	533	17	AQ580338	AQ580338 T135971b
c	35	65.4	6.2	473	17	AQ866027	AQ866027 nbeb0026P
c	36	65.4	6.2	901	17	AQ867999	AQ867999 nbeb0018H
c	37	62	5.9	342	17	AQ841745	AQ841745 T132158b
c	38	62	5.9	457	17	BH168898	BH168898 SALK_0003
	39	57.2	5.5	447	13	BI420478	BI420478 LjNEST57f
	40	56.6	5.4	586	13	BM191106	BM191106 daj32b10.
c	41	55.6	5.3	569	17	AQ580359	AQ580359 T136000b
	42	50.2	4.8	151	10	BE022923	BE022923 sm89c11.y
c	43	47	4.5	711	13	BJ169253	BJ169253 BJ169253
	44	46.6	4.4	620	9	AL647838	AL647838 AL647838
	45	45.8	4.4	161	17	B24158	B24158 F19B10TR IG

ALIGNMENTS

RESULT 1
AU226180/c
LOCUS AU226180 435 bp mRNA linear EST 23-APR-2002
DEFINITION AU226180 RAFL14 Arabidopsis thaliana cDNA clone RAFL14-14-A20 3', mRNA sequence.
ACCESSION AU226180
VERSION AU226180.1 GI:19740827
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 435)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Akiyama,K., Enju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998).cDNA cleaved with BamHI

and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="RAFL14-14-A20"
                     /clone_lib="RAFL14"
                     /tissue_type="root"
                     /lab_host="DH10B"
                     /note="Site_1: BamHI; Site_2: SalI"
BASE COUNT      117 a     85 c     85 g    148 t
ORIGIN

  Query Match             27.5%;  Score 288.8;  DB 9;  Length 435;
  Best Local Similarity   90.3%;  Pred. No. 2.2e-61;
  Matches  327;  Conservative    0;  Mismatches   17;  Indels   18;  Gaps    1;

Qy        706 GAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGA 765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        435 GAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCTTGA 376

Qy        766 AGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTAC 825
              |||||||||||||||||||||||||||||| || ||||||| ||||||||||||||||
Db        375 AGCCAAACAGAATCAGGCTTGGGAACTGGGAGGTTCATCCTCTATCAAAGCAGCAGTTAC 316

Qy        826 AATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTC 885
              |||||||||||||||||||||||||||||||||| || ||||||||||||||||||||
Db        315 AATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTCACCAGGTTCTTAAGGACCTTC 256

Qy        886 CTGATGCTGTCAGTGGCTCATAACGTGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCC 945
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        255 CTGATGCTGTCAGTGGCTCATAACGTAAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCC 196

Qy        946 AAGAGTTAGCATCAAATGATATGATACACCTAATCTAGTCAAGTAGATGCAATTCTT--- 1002
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 AAGAGTTAGCATCAAATGATATGATACACCTAATCTAGTCAAGTAGATGCAATTCTTATG 136

Qy       1003 --------------GTGAATATTGTATCTAGTTCTGGTCCCTTTAACCGTCCAGAAACT 1047
                             |||| |  |||| |||||  ||||| ||||| | ||||||||||
Db        135 GGATTTACAAAGATATTGAACACAGTATTTAGTTGCGGTCCATTTAATCTTCCAGAAACT 76

Qy       1048 AG 1049
              ||
Db         75 AG 74
```

```
RESULT 2
BE659005/c
LOCUS       BE659005                840 bp    mRNA    linear   EST 24-MAY-2001
DEFINITION  GM700008A10G11 Gm-r1070 Glycine max cDNA clone Gm-r1070-2997 3',
            mRNA sequence.
ACCESSION   BE659005
VERSION     BE659005.1  GI:9984897
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 840)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: AW460204 corresponding to Gm-c1015-3787 (5')
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1070-2997"
                     /clone_lib="Gm-r1070"
                     /note="The library Gm-r1070 is a sequence-driven, reracked
                     set of 9,216  clones selected from cDNA libraries from
                     various tissues and stages  of development of soybean that
                     represent 2,639 sequences from  immature cotyledons, 1,770
                     from immature seed coats, 3,938 from  flowers, and 869
                     from young pods.  The 5' ESTs of the source clones from
                     the different libraries was  used to select singletons, or
                     a representative of each contig, which  were reracked to
                     form library Gm-r1070.  The cDNA clones of the reracked
                     Gm-r1070 library were then sequenced  at the 3' end.  The
                     contig analysis to select unique genes was performed by
                     the  laboratory of Ernest Retzel, Center for Computational
                     Genomics and  Bioinformatics, University of Minnesota,
                     http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
                     .  Reracking was performed by Genome Systems, St. Louis,
                     http://www.genomesystems.com, and 3' sequencing by the
                     Keck Center  for Comparative and Functional Genomics,
                     University of Illinois,
                     http://www.life.uiuc.edu/biotech/keck.html.  Note: The
                     corresponding 5' EST from each clone in the Gm-r1070
                     library is listed in the 'OTHER EST' field.  The detailed
                     information on the source library for each clone can  also
                     be obtained by referring to the Genome Systems clone ID of
                     the  original cDNA library that is also listed under
                     'OTHER EST'."
BASE COUNT      242 a    167 c    158 g    240 t     33 others
ORIGIN

  Query Match             26.1%;  Score 273.8;  DB 10;  Length 840;
  Best Local Similarity   66.1%;  Pred. No. 1.3e-57;
  Matches  392;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy        300 AGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAG 359
              | ||  || || | ||||| |||| |||||   ||||  ||||||| ||||  |  ||
Db        727 ATGACATTGCCTGTAATGAAGTTTAGTGGTCAAATTTCTTATAGCAGGACTTTTGATGCT 668

Qy        360 GTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGA 419
              || || ||   ||||  |||  |  || ||||  |    || |    ||
Db        667 GTAGAGAAAGCTGCAACAAAGCNCTTACAAATTCTCCAAGAAAAAACGACCGACATGATG 608

Qy        420 ATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCG 479
                | ||    |||| || || |||||||||| |||     | ||||||||||||  ||
Db        607 CAAACTGCAATTGGATTTGACATTGAGTGGAAACCCNNCNTCAGAAAAGGTGTTCCTCCC 548

Qy        480 GGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATT 539
              || ||||| ||      | ||||||||||  || | ||   ||||| ||||| | ||| |
Db        547 GGAAAGGTAGCAGTGATGCAGATATGTGGTGACACTAGACATTGTCATGTTCTACATCTA 488

Qy        540 TTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAG 599
              |||||||||| |||||||||| | | || | | |  |||||||| | ||| |  | |||
Db        487 ATTCATTCTGGAATCCCTCAAAATTTACAGCTTTTGCTTGAAGATCCCACAGTCTTGAAG 428

Qy        600 GTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATC 659
              || ||   ||| |||||||||||  |||||||| ||||    || |||    |  | |
Db        427 GTTGGAGCTGGGATTGATGGTGATGCTGTGAAGGTTTTTAGAGATTATAACATATCTGTT 368
```

```
Qy        660 AAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGG 719
              |||| |||   |||||||||   |   || || |||||  ||||||||||| | || |||
Db        367 AAAGGTGTGACGGATCTTTCTTTTCATGCTAATCAAAAGCTTGGTGGAGATCATAAGTGG 308

Qy        720 GGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATC 779
              || ||||| ||  | ||||| | |||| | |  ||| ||||    |||||||||| | ||
Db        307 GGTCTTGCATCTTTGACTGAAAAACTTCTATCAAAACAGCTTAAAAAGCCCAACAAAATA 248

Qy        780 AGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACG 839
              || || || ||||||  |  |  | |||||||| ||||  || | || || |||||
Db        247 AGACTGGGAAATTGGGAGGCTCCTGTTTTGTCAAAGGAGCAACTAGAGTATGCTGCAACA 188

Qy        840 GATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGC 892
              ||||||| |||||| |||  ||||||  |||   |||| || || || || ||
Db        187 GATGCTTTTGCTTCTTGGTGTCTTTATCAGGCGATTAAAGATCTCCCGGACGC 135


RESULT 3
BG441595
LOCUS       BG441595                704 bp    mRNA    linear   EST 15-MAR-2001
DEFINITION  GA__Ea0013O22f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ea0013O22f, mRNA sequence.
ACCESSION   BG441595
VERSION     BG441595.1  GI:13351247
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 704)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 701.
FEATURES             Location/Qualifiers
     source          1. .704
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0013O22f"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      188 a    144 c    155 g    217 t
ORIGIN

  Query Match             24.9%;  Score 261;  DB 12;  Length 704;
  Best Local Similarity   63.7%;  Pred. No. 1.9e-54;
  Matches  434;  Conservative    0;  Mismatches  235;  Indels   12;  Gaps    2;

Qy        145 CTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATC 204
              || |  || || ||||  |    |   |  |  |   |    |  || |    |  |  |
Db         24 CTCCTCCTACTCCTCCTTCTCCTCCTTCATCTTCTCGCAAGAAACGCAAGTCCGCCGTCC 83

Qy        205 CAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACAT 264
                |||||     |   || | | | |||||||||||| | ||| ||    ||  |  |
Db         84 ATCATCAAGATAACCTCAAAACCAGCCGCCAATTGCCCCATTCTATTCTCTCACCCCCGC 143

Qy        265 CTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA---------GGAATTTTCCAGCAA 315
              ||| |   |  | | |  || | ||||| | | |||           |  | |||  |
Db        144 CTTCTTTTCCGTCTTCCTTCGCTCGTTGTCAATCTAATACAAGGCTTAGATATCCGCCGT 203

Qy        316 TGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCGAGCAA 375
              |||| ||||| |||   ||||||||||||   || |    |||||| || ||    ||||
Db        204 TGAGATTTGGAGGTCATATTTTGTATAGCTTTACGGAAGATGAGGTAGAAAATGCTGCAA 263

Qy        376 TGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCT 435
              || ||||  | ||| || |||| |  ||||   | |||   ||   || |  | | || |
Db        264 TGGAGCTATTAAAAATTGTTGAAATTAAGAAGAAAGAAATGGGTCAAGTTGCTTTAGGGT 323

Qy        436 TGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTG 495
              | ||||||||||||||  ||   |||   ||||||  || | || |||||||  ||||
Db        324 TTGATATTGAGTGGAAGCCCTCTTTCCAAAAAGGAATTTTGCCTGGGAAGGCTGCGGTGA 383

Qy        496 TCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCC 555
              | |||||||||   || ||| | |||||| |||| ||||||||||||||||||||| || |
Db        384 TGCAGATATGTTGTGACAGTCAGTATTGTTATGTGATGCATATTTTTCATTCTGGCATAC 443

Qy        556 CTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTG 615
              |||||||||| ||   |||| |||| || ||   | || |||| || ||  ||| |||||
Db        444 CTCAAAGTCTGCAGGTTCTTCTTGAGGACTCCGAAATTATAAAAGTTGGAGTTGCAATTG 503

Qy        616 ATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATC 675
              |||| ||  |||| ||| | |||  |||||||  |||   | | || | | | || ||||
Db        504 ATGGCGATGCTGTCAAGGTGTTCAGTGACTATAAAGTATCTGTTAATGCTTTGGAAGATC 563

Qy        676 TTTCAGATTTAGCCAACCAAAAAATTGGTGGAGAT---AAAAAATGGGGCCTTGCCTCAC 732
              |||||||| |||| || |||||| ||||| |  ||        | ||| | |||||  | |
Db        564 TTTCAGATCTAGCGAATCAAAAATTTGGTAGTAATTGCCGGCACTGGAGTCTTGCTGCTC 623

Qy        733 TAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACT 792
              | || |||    || |||||||||||||| | |||||| || | ||| ||||| || ||||
Db        624 TGACAGAGGAGTTTATTTGCAAAGAGCTTCCGAAGCCCAAGAAAATTAGGCTGGGAAACT 683

Qy        793 GGGAGTTTTATCCTCTGTCAA 813
              ||||||| |||||| | || |
Db        684 GGGAGTTATATCCTTTATCCA 704


RESULT 4
AW201789
LOCUS       AW201789                552 bp    mRNA    linear   EST 02-DEC-2001
DEFINITION  sf07c10.y1 Gm-c1027 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1027-1579 5' similar to TR:Q9Z242 Q9Z242 WRN PROTEIN. ;, mRNA
            sequence.
ACCESSION   AW201789
VERSION     AW201789.1  GI:6482532
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
```

South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccu@resgen.com
High quality sequence stop: 410.

```
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1027-1579"
                     /clone_lib="Gm-c1027"
                     /tissue_type="cotyledons of 3- and 7-day-old Williams
                     seedlings"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from cotyledons of 3- and 7-day-old Williams seedlings
                     which were propagated on paper towels with distilled
                     water.  The cotyledons were flash-frozen in liquid
                     nitrogen, then lyophilized for 72 hours.  Unequal amounts
                     of mRNA was used for cDNA synthesis.  Stratagene's cDNA
                     Synthexix Kit (catalog number 200401) was used to
                     synthesize the cDNA.  First- stranded synthesis was
                     performed with 5-methyl dCTP, hence the ligated cDNA was
                     hemimethylated.  A modification of Stratagene's
                     first-strand  synthesis primer was used.  An anchor
                     nucleotide (V=A, C, or G) was added to the 3' end of the
                     primer [GAGAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)18] to anchor
                     the primer at the 5' end of the poly(A) tract.  After
                     second- strand synthesis, the cDNA ends were filled in
                     with cloned Pfu DNA, ligated to EcoRI adapters and
                     subsequently phosphorylated.  The XhoI site within the
                     first-strand synthesis primer was then restricted by
                     digestion with XhoI; all XhoI sites in the cDNA would be
                     protected by their hemimethylated status.  The cDNA
                     constructs were size-fractionated with a 500 bp cutoff,
                     using GibcoBRL Life Technologies' cDNA Size Fractionation
                     column.  The column eluent was then ligated into
                     Stratagene's pBluescript(tm) II XR Predigested vector
                     (pBluescript II SK(+) that has been digested with EcoRI
                     and XhoI, and phosphorylated by Stratagene).  97% of the
                     white and blue colonies appear to contain recombinant
                     plasmids with cDNA inserts, based on size (n=30). This
                     library was constructed by Dr. Paul Keim and Dr. Virginia
                     Coryell."
BASE COUNT      168 a    103 c    125 g    154 t      2 others
ORIGIN
```

Query Match 23.4%; Score 245; DB 10; Length 552;
Best Local Similarity 66.5%; Pred. No. 1.9e-50;
Matches 350; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

```
Qy        367 AGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTT 426
Db          4 AGGCTGCAACAAAGCTCTTACAAATTCTCCAAGAAAAAACGACCGACATGATGCAAACTG 63

Qy        427 TTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGG 486
Db         64 CAATTGGATTTGACATTGAGTGGAAACCCACCTTCAGAAAAGGTGTTCCTCCCGGAAAGG 123

Qy        487 TTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATT 546
Db        124 TAGCAGTGATGCAGATATGTGGTGACACTAGACATTGTCATGTTCTACATCTAATTCATT 183

Qy        547 CTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTA 606
Db        184 CTGGAATCCCTCAAAATTTACAGCTTTTGCTTGAAGATCCCACAGTCTTGAAGGTTGGAG 243

Qy        607 TTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATG 666
Db        244 CTGGGATTGATGGTGATGCTGTGAAGGTTTTTAGAGATTATAACATATCTGTTAAAGGTG 303

Qy        667 TTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTG 726
Db        304 TGACGGATCTTTCTTTTCATGCTAATCAAAAGCTTGGTGGAGATCATAAGTGGGGTCTTG 363

Qy        727 CCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTG 786
Db        364 CATCTTTGACTGAAAAACTTCTATCAAAACAGCTTAAAAAGCCCAACAAAATAAGACTGG 423

Qy        787 GGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTT 846
Db        424 GAAATTGNGAGGCTCCTGTTTTGTCAAAGGAGCAACTAGAGTATGCTGCAACAGATGCTT 483

Qy        847 ATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGC 892
Db        484 TTGCTTCTTGGTGTCTTTATCAGGCGATTAAAGATCTCCCGGACGC 529
```

```
RESULT 5
AU237507
LOCUS       AU237507                671 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AU237507 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-31-E24 5',
            mRNA sequence.
ACCESSION   AU237507
VERSION     AU237507.1  GI:19876676
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Akiyama,K., Enju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J.,
            Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu
            ,M., Hayashizaki,Y. and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998).cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .671
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="RAFL16-31-E24"
                     /clone_lib="RAFL16"
                     /lab_host="DH10B"
                     /note="Site_1: BamHI; Site_2: SalI; dark-grown"
BASE COUNT      171 a    109 c    164 g    225 t      2 others
ORIGIN
```

Query Match 23.2%; Score 243.8; DB 9; Length 671;
Best Local Similarity 94.9%; Pred. No. 3.8e-50;
Matches 263; Conservative 0; Mismatches 9; Indels 5; Gaps 1;

```
Qy        561 AGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGT 620
Db        395 AGACTCCAACATCTTATTGAGGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGT 454

Qy        621 GACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGT-----ATCAAAGATGTTGAGGATC 675
Db        455 GACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAATCAAAGATGTTGAGGATC 514

Qy        676 TTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAA 735
```

```
Db        515 TTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAA 574

Qy        736 CTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGG 795
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        575 CTGAGACACTTGTTTGCAAAGAGCTCTTGAAGCCAAACAGAATCAGGCTTGGGAACTGGG 634

Qy        796 AGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGC 832
              ||  | || |||| ||||||||||||||||||| |||
Db        635 AGGGTCATNCTCTATCAAAGCAGCAGTTACAATNCGC 671
```

RESULT 6
AW832139

```
LOCUS       AW832139                467 bp    mRNA    linear   EST 03-DEC-2001
DEFINITION  sm20b11.y1 Gm-c1027 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1027-9430 5' similar to SW:WRN_HUMAN Q14191 WERNER SYNDROME
            HELICASE. [1] ;, mRNA sequence.
ACCESSION   AW832139
VERSION     AW832139.1  GI:7926113
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 734   Std Error: 0.00
            High quality sequence stop: 417.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1027-9430"
                     /clone_lib="Gm-c1027"
                     /tissue_type="cotyledons of 3- and 7-day-old Williams
                     seedlings"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from cotyledons of 3- and 7-day-old Williams seedlings
                     which were propagated on paper towels with distilled
                     water.  The cotyledons were flash-frozen in liquid
                     nitrogen, then lyophilized for 72 hours.  Unequal amounts
                     of mRNA was used for cDNA synthesis.  Stratagene's cDNA
                     Synthexix Kit (catalog number 200401) was used to
                     synthesize the cDNA.  First- stranded synthesis was
                     performed with 5-methyl dCTP, hence the ligated cDNA was
                     hemimethylated.  A modification of Stratagene's
                     first-strand  synthesis primer was used.  An anchor
                     nucleotide (V=A, C, or G) was added to the 3' end of the
                     primer [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)18] to anchor
                     the primer at the 5' end of the poly(A) tract.  After
                     second- strand synthesis, the cDNA ends were filled in
                     with cloned Pfu DNA, ligated to EcoRI adapters and
                     subsequently phosphorylated.  The XhoI site within the
                     first-strand synthesis primer was then restricted by
                     digestion with XhoI; all XhoI sites in the cDNA would be
                     protected by their hemimethylated status.  The cDNA
                     constructs were size-fractionated with a 500 bp cutoff,
                     using GibcoBRL Life Technologies' cDNA Size Fractionation
                     column.  The column eluent was then ligated into
                     Stratagene's pBluescript(tm) II XR Predigested vector
                     (pBluescript II SK(+) that has been digested with EcoRI
                     and XhoI, and phosphorylated by Stratagene).  97% of the
                     white and blue colonies appear to contain recombinant
                     plasmids with cDNA inserts, based on size (n=30). This
                     library was constructed by Dr. Paul Keim and Dr. Virginia
                     Coryell."
BASE COUNT      133 a     84 c    109 g    140 t      1 others
ORIGIN
```

```
  Query Match             23.0%;  Score 241.2;  DB 10;  Length 467;
  Best Local Similarity   70.0%;  Pred. No. 1.6e-49;
  Matches  324;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy        430 TTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTG 489
              |||| || || |||||||||| ||| |  || ||||||||||||||  || || ||||| |
Db          4 TTGGATTTGACATTGAGTGGAAACCCACCTTCAGAAAAGGTGTTCCTCCCGGAAAGGTAG 63

Qy        490 CGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTG 549
              |    | ||||||||||  || | ||   ||||| ||||| | ||| |  |||||||||
Db         64 CAGTGATGCAGATATGTGGTGACACTAGACATTGTCATGTTCTACATCTAATTCATTCTG 123

Qy        550 GTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTG 609
              | |||||||||| | | || | | |  |||||||| | ||| |  | ||||| ||   ||
Db        124 GAATCCCTCAAAATTTACAGCTTTTGCTTGAAGATCCCACAGTCTTGAAGGTTGGAGCTG 183

Qy        610 GAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTG 669
              | ||||||||||  |||||||| ||||    || |||    |   | | |||| |||
Db        184 GGATTGATGGTGATGCTGTGAAGGTTTTTAGAGATTATAACATATCTGTTAAAGGTGTGA 243

Qy        670 AGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCT 729
              |||||||||   |   || || |||||  ||||||||||| | || ||||| ||||| |
Db        244 CGGATCTTTCTTTTCATGCTAATCAAAAGCTTGGTGGAGATCATAAGTGGGGTCTTGCAT 303

Qy        730 CACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGA 789
              |  | ||||| | |||| | |  ||| ||||    ||||| |||| ||| || || || |
Db        304 CTTTGACTGAAAAACTTCTATCAAAACAGCTTAAAAAGCCCAACAAAATAAGACTGGGAA 363

Qy        790 ACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATG 849
              | ||||||  |  |  | |||||||| |||| || | || || ||||| |||||||
Db        364 ATTGGGAGGCTCCTGTTTTGTCAAAGGAGCAACTAGAGTATGCTGCAACAGATGCTTTTG 423

Qy        850 CTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGC 892
              |||| |||  ||| ||  |||   |||| || || || || ||
Db        424 CTTCTTGGTGTCTNTATCAGGCGATTAAAGATCTCCCGGACGC 466
```

RESULT 7
BQ799068

```
LOCUS       BQ799068                661 bp    mRNA    linear   EST 30-JUL-2002
DEFINITION  EST 1237 Green Grape berries Lambda Zap II Library Vitis vinifera
            cDNA clone GT172E01 3', mRNA sequence.
ACCESSION   BQ799068
VERSION     BQ799068.1  GI:22014034
KEYWORDS    EST.
SOURCE      Vitis vinifera.
  ORGANISM  Vitis vinifera
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Vitaceae; Vitis.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   Abbal,P., Agasse,A., Ageorges,A., Atanassova,R., Barrieu,F.,
            Couture,C., Dedaldechamp,F., Delrot,S., Glissant,D., Grimplet,J.,
            Hamdi,S., Romieu,C. and Terrier,N.
  TITLE     Generation of Expressed Sequence Tag from Grape Berry (skin, pulp
            or seeds) at Various Developmental Stages
```

```
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Romieu C.
            Unite de Recherche des Produits de la Vigne
            Institut National de la Recherche Agronomique
            2, place Viala, 34 060 Montpellier Cedex 01, France
            Tel: 00-33-(0)4-99-61-28-62
            Fax: 00-33-(0)4-99-61-28-57
            Email: romieu@ensam.inra.fr
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .661
                     /organism="Vitis vinifera"
                     /cultivar="Shiraz"
                     /db_xref="taxon:29760"
                     /clone="GT172E01"
                     /clone_lib="Green Grape berries Lambda Zap II Library"
                     /dev_stage="green stage"
                     /note="Organ: Fruit; Vector: Lambda Zap II; Site_1: Eco RI
                     ; Site_2: XhoI; Oriented library, construction described
                     in Generation of ESTs from grape Berry (skin, pulp or
                     seeds) at various developmental stages by Terrier,N.,
                     Ageorges,A., Abbal,P., Romieu,C. in J. Plant Physiol. 158
                     (12): 1575-83 2001"
BASE COUNT      191 a    130 c    156 g    184 t
ORIGIN

  Query Match           22.4%;  Score 235.4;  DB 14;  Length 661;
  Best Local Similarity   69.8%;  Pred. No. 4.7e-48;
  Matches  333;  Conservative    0;  Mismatches  141;  Indels    3;  Gaps    1;

Qy        423 GCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGG 482
              ||  || |||| || || ||||||||||||||| |  ||||||||||||||    | |||
Db         11 GCCATTCTTGGATTTGACATTGAGTGGAGACCCACCTTTAGAAAAGGTGTGTCACAAGGG 70

Qy        483 AAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTT 542
              ||||  |||   | | ||||||||||  | || || | ||||| | || |||||||| ||
Db         71 AAGGCTGCAGTTATGCAGATATGTGGGGGTAATAGTCATTGTTACGTCATGCATATTATT 130

Qy        543 CATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTA 602
              |||||||| ||||| | ||  || |||  |||  |||| ||| | |||    |||||||
Db        131 CATTCTGGAATCCCGCGAAACCTGCAATCTCTACTTGAGGATCCTACATCCATAAAGGTT 190

Qy        603 GGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAA 662
              ||  |||| ||||    |||  |||| ||| ||||| | ||  || | ||   | | |||
Db        191 GGAGTTGGCATTGCCAATGATGCTGTTAAGGTTTTCAAAGATCATAGTGTATCTGTTAAA 250

Qy        663 GATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGA---TAAAAAATGG 719
              ||| | || ||||| ||  || | || || |||||  ||||||||||    || || |||
Db        251 GATTTGGAAGATCTATCCTATCTTGCTAATCAAAAGCTTGGTGGAGAAGCAAAGAAGTGG 310

Qy        720 GGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATC 779
               | ||||  || ||||| || |  ||| ||| |||| |||| || ||||| |||| |||
Db        311 AGTCTTGGATCTCTAACGGAAATGCTTATTTCCAAACAGCTTCTTAAGCCCAACAAAATA 370

Qy        780 AGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACG 839
              ||  | || |||||||||  | ||  | | ||||||  ||  || | || || || ||
Db        371 AGATTGGGAAACTGGGAGGCTGATGTTTTATCAAAGGCCCAACTAGAGTATGCTGCTACA 430

Qy        840 GATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTC 896
              ||||||| ||| |||||| | || ||| |||| || |||  | | || ||| | | |
Db        431 GATGCTTTTGCCTCATGGTACCTATACGAGGTACTAAAGAGCTTCCCAGATACAGCC 487


RESULT 8
BM406124
LOCUS       BM406124                 535 bp    mRNA    linear   EST 22-JAN-2002
DEFINITION  EST580451 potato roots Solanum tuberosum cDNA clone cPRO26F14 5'
            end, mRNA sequence.
ACCESSION   BM406124
VERSION     BM406124.1  GI:18257754
KEYWORDS    EST.
SOURCE      potato.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   van der Hoeven,R., Sun,H., Karamycheva,S.A., Tsai,J., Van Aken,S.,
            Utterback,T., Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C.,
            Tanksley,S. and Baker,B.
  TITLE     Generation of ESTs from potato roots
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Research Genetics, Libraries Division
            Tel: 1-800-711-6195
            Email: cdna@resgen.com
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="cPRO26F14"
                     /clone_lib="potato roots"
                     /tissue_type="roots"
                     /dev_stage="in vitro grown stem cuttings"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; supplier: Cornell University, Tanksley lab;
                     sequencing; The Institute for Genomic Research. Roots were
                     isolated from in vitro grown stem cuttings on CM medium.
                     Roots were isolated two weeks after placing the stem
                     cuttings from in vitro grown plants on medium."
BASE COUNT      146 a    101 c    136 g    152 t
ORIGIN

  Query Match           22.3%;  Score 233.8;  DB 13;  Length 535;
  Best Local Similarity   64.9%;  Pred. No. 1.1e-47;
  Matches  346;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy        363 GATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATA 422
              || ||    ||| |  |||| | ||  || ||||    ||||       ||   || |
Db          2 GAGAAATCTGCAGTAGAGCTGTTAAATTTTGTTGAGGAAAAGAAGAGGAAAGAAGGGAAT 61

Qy        423 GCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGG 482
              | |  | |||| || || ||||| ||||  ||   |||||||| |||||| |  || |||
Db         62 GTTGCTCTTGGATTTGACATTGAATGGAAGCCCTCTTTTAGAAGAGGTGTGCCACCTGGG 121

Qy        483 AAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTT 542
              ||||  ||   | | ||||||||||  || |    | ||||| ||||| |||||||  |
Db        122 AAGGCCGCTGTTATGCAGATATGTGGTGACAAGGGTAATTGTTATGTTCTGCATATCATC 181

Qy        543 CATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTA 602
              || ||||| |||||||||| ||| |||  |||| |||| ||| ||||  |||| |||||
Db        182 CACTCTGGAATCCCTCAAACTCTTCAATCTCTTCTTGAGGATCCAACTGTTGTGAAGGTG 241

Qy        603 GGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAA 662
              ||| || | ||||    |||  ||   ||  ||  ||| ||  |    ||   | | ||
Db        242 GGTGTTTGCATTGCAAATGATGCTTACAAAGTTCGCCAAGATCACAATGTATCTGTGAAG 301

Qy        663 GATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGC 722
              | | | || |||||||| ||  | ||||| ||||  ||| ||     || || ||| |
Db        302 GCTTTGGAAGATCTTTCTGAACTTGCCAACAAAAAGCTTGATGATCCCAAGAAGTGGAGT 361

Qy        723 CTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGG 782
              || || |||||||| |||  ||| ||  |||  |||||  |||||| ||  ||||||
Db        362 CTAGCATCACTAACCGAGAAGCTTCTTGCCAAGCAGCTCCCTAAGCCAAGCAATATCAGG 421

Qy        783 CTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGAT 842
              | ||| | ||||||  | ||    | || | | | ||  |||| || || || || |||
Db        422 TTGGGGGATTGGGAGGCTAATGTGTTATCTAGGGACCAACTACATTATGCTGCTACAGAT 481

Qy        843 GCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGT 895
              || | || ||| |||||  | ||  |||  || |||  |||||| ||   |||
Db        482 GCCTTTGTTTCCTGGCACTTGTATCAGGCACTGAAGAGCCTTCCGGAGATTGT 534
```

```
RESULT 9
BG133220
LOCUS       BG133220                506 bp    mRNA    linear   EST 31-JAN-2001
DEFINITION  EST466112 tomato crown gall Lycopersicon esculentum cDNA clone
            cTOE11B12 5' sequence, mRNA sequence.
ACCESSION   BG133220
VERSION     BG133220.1  GI:12633408
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Hansen,C., Ronning
            ,C. and Tanksley,S.
  TITLE     Generation of ESTs from tomato crown gall tissue
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cTOE11B12"
                     /clone_lib="tomato crown gall"
                     /tissue_type="crown gall"
                     /dev_stage="crown galls from full-grown plants (8 wks old
                     )"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2:
                     Xhol; Four wk old greenhouse plants were stab inoculated
                     on stem with Agrobacterium tumefaciens C58 (Dr. T.J. Burr,
                     Cornell U.). Galls were allowed to develop for another 4
                     wks, when gall tissue was frozen in  liquid nitrogen."
BASE COUNT      142 a     91 c    126 g    147 t
ORIGIN

  Query Match             21.3%;  Score 223;  DB 12;  Length 506;
  Best Local Similarity   65.2%;  Pred. No. 5.6e-45;
  Matches  328;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy        363 GATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATA 422
              || ||    ||| |  |||| | ||  || ||||    ||||      ||   || |
Db          2 GAGAAATCTGCAGTAGAGCTGTTAAATTTTGTTGAGGAAAAGAAGAGGAAAGAAGGGAAT 61

Qy        423 GCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGG 482
              | |  | |||| || || ||||| |||| || | |||||||| |||||| |  || |||
Db         62 GTTGCTCTTGGATTTGACATTGAATGGAAGCCCACTTTTAGAAGAGGTGTGCCACCTGGG 121

Qy        483 AAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTT 542
              |||| |||   | | |||||||||||  || |    | ||||| ||||| |||||||  |
Db        122 AAGGCTGCTGTTATGCAGATATGTGGTGACAAGGGTAATTGTTATGTTTTGCATATCATC 181

Qy        543 CATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTA 602
              || || || |||||||||| ||| |||  |||| |||| ||| ||||  |||| |||||
Db        182 CACTCCGGAATCCCTCAAACTCTGCAATCTCTTCTTGAGGATCCAACTGTTGTGAAGGTG 241

Qy        603 GGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAA 662
              ||| || | ||||    |||  || | ||  ||  ||| ||  |    ||   | | ||
Db        242 GGTGTTTGCATTGCAAATGATGCTTTCAAAGTTCGCCAAGATCACAATGTATCTGTGAAG 301

Qy        663 GATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGC 722
              | | | || |||||||| ||  | |||||| ||||  ||| ||      || || ||| |
Db        302 GCTTTGGAAGATCTTTCCGAACTTGCCAACAAAAAGCTTGATGATCCCAAGAAGTGGAGT 361

Qy        723 CTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGG 782
              || || ||| |||||||||  ||| ||  |||  ||||||  |||||||  |  ||||||
Db        362 CTAGCATCATTAACTGAGAAGCTTCTTGCCAAGCAGCTCCCAAAGCCAAGTAATATCAGG 421

Qy        783 CTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGAT 842
              | ||||| ||||||  | ||    | || | | | ||  |||| || || || || |||
Db        422 TTGGGGAATTGGGAGGCTAATGTGTTATCTAGGGACCAACTACATTATGCTGCTACAGAT 481

Qy        843 GCTTATGCTTCATGGCATCTTTA 865
              || | || ||| ||| |  | ||
Db        482 GCCTTTGTTTCCTGGTACTTATA 504


RESULT 10
BM094501
LOCUS       BM094501                531 bp    mRNA    linear   EST 30-NOV-2001
DEFINITION  saj16g03.y1 Gm-c1066 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1066-3005 5' similar to TR:Q9SVM6 Q9SVM6 HYPOTHETICAL 35.1
            KD PROTEIN. ;, mRNA sequence.
ACCESSION   BM094501
VERSION     BM094501.1  GI:17023467
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1066-3005"
                     /clone_lib="Gm-c1066"
                     /tissue_type="Leaf and shoot tip, salt stressed, 2 week
                     old seedling"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from unexpanded leaves and the shoot tips of 2 week old
                     seedling from the cultivar Williams. The 2 week old
                     seedlings were salt stressed in a solution of 500mM NaCl
                     for 3 days prior to harvesting. Complementary DNA was
                     synthesized from mRNA using a  primer consisting of a
                     poly(dT) sequence with a XhoI restriction site. EcoRI
                     adapters were ligated to the blunt-ended cDNA fragments
                     followed by XhoI  digestion. The cDNA fragments were
                     directionally cloned into the EcoRI-XhoI  restriction site
                     of the pBluescript vector. The ligated cDNA fragments were
                     transformed into DH10B host cells (GibcoBRL). This library
                     was constructed  in the laboratory of Dr. Randy
                     Shoemaker."
BASE COUNT      164 a     85 c    119 g    163 t
```

```
ORIGIN

  Query Match            20.4%;  Score 213.6;  DB 13;  Length 531;
  Best Local Similarity  64.6%;  Pred. No. 1.2e-42;
  Matches  340;  Conservative    0;  Mismatches  174;  Indels   12;  Gaps    1;

Qy        299 TAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGA 358
              || ||  || ||   ||||| |||| |||||   ||||| |||| || ||||  |  ||
Db         18 TATGAGATTGCCTCTAATGAAGTTTAGTGGTCAAATTTTTTATATCAGGACTTTTGATGC 77

Qy        359 GGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGG 418
              || || ||    ||||   ||||  |  || ||||  |    | ||   |
Db         78 TGTAGAGAAAGCTGCAACAAAGCTCTTACAAATTCTCCAAGAGATGATGCA--------- 128

Qy        419 AATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCC 478
              |  |    |||| || || |||||||||| ||| |  || ||||||||||||  ||
Db        129 ---AATTGCAATTGGATTTGACATTGAGTGGAAACCCACCTTCAGAAAAGGTGTTCCACC 185

Qy        479 GGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATAT 538
              || ||||| ||     | |||||||| ||  || | ||   ||||| ||||| | ||| |
Db        186 CGGAAAGGTAGCAGTGATGCAGATATATGGTGACACTAGACATTGTCATGTTCTACATCT 245

Qy        539 TTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAA 598
              ||||||||||| ||||||  || | | || | | |  |||||||| | ||| |  | ||
Db        246 AATTCATTCTGGAATCCCTTGAAATTTACAGCTTTTGCTTGAAGATCCCACAGTCTTGAA 305

Qy        599 GGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTAT 658
              ||| ||   ||| |||||||||||  |||||||| ||||    || |||    |   | |
Db        306 GGTTGGAGCTGGGATTGATGGTGATGCTGTGAAGGTTTTTAGAGATTATAACATATCTGT 365

Qy        659 CAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATG 718
              ||||  |||   |||||||||    |   || || | |||  |||||||||| | || ||
Db        366 TAAAGGTGTGACGGATCTTTCTTTTCATGCTAATCGAAAGCTTGGTGGAGATCATAAGTG 425

Qy        719 GGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAAT 778
              ||| ||||| ||  | ||||| | |||| | |  ||| ||||    |||| |||| |||
Db        426 GGGTCTTGCATCTTTGACTGAAAAACTTTTATCAAAACAGCTTAAAAAGCCTAACAAAAT 485

Qy        779 CAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTA 824
              || || || || ||||||  |  |  | |||||||| ||||  ||
Db        486 AAGACTGGGAAATTGGGAGACTCCTGTTTTGTCAAAGGAGCAACTA 531


RESULT 11
BM358892
LOCUS       BM358892                515 bp    mRNA    linear   EST 09-JAN-2002
DEFINITION  GA__Ea0013P02r Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ea0013P02r, mRNA sequence.
ACCESSION   BM358892
VERSION     BM358892.1  GI:18099638
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total High Quality bases = 415
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 514.
FEATURES             Location/Qualifiers
     source          1. .515
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0013P02r"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      134 a    112 c    109 g    160 t
ORIGIN

  Query Match            16.4%;  Score 172.2;  DB 13;  Length 515;
  Best Local Similarity  64.6%;  Pred. No. 2.6e-32;
  Matches  277;  Conservative    0;  Mismatches  143;  Indels    9;  Gaps    1;

Qy        208 ATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTT 267
              |||||     |   || | | | |||||||||| | ||| ||    ||  |  | |||
Db         87 ATCAAGATAACCTCAAAACCAGCCGCCAATTGCCCCATTCTATTCTCTCACCCCCGCCTT 146

Qy        268 ATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA---------GGAATTTTCCAGCAATGA 318
              |   |  | | | |  || | ||||| | | |||           |  | |||  |  |||
Db        147 CTTTTCCGTCTTCCTTCGCTCGTTGTCAATCTAATACAAGGCTTAGATATCCGCCGTTGA 206

Qy        319 GGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGC 378
              | ||||| |||   |||||||||||   || |    ||||| || ||    ||||||
Db        207 GATTTGGAGGTCATATTTTGTATAGCTTTACGGAAGATGAGGTAGAAAATGCTGCAATGG 266

Qy        379 AGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGG 438
              ||||  | ||| || |||| |  ||||   | |||   ||   || |  | | || || |
Db        267 AGCTATTAAAAATTGTTGAAATTAAGAAGAAAGAAATGGGTCAAGTTGCTTTAGGGTTTG 326

Qy        439 ATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCC 498
              ||||||||||||  ||   |||   ||||||  || | ||  ||||||| ||||    | |
Db        327 ATATTGAGTGGAAGCCCTCTTTCCAAAAAGGAATTTTGCCTGGGAAGGCTGCGGTGATGC 386

Qy        499 AGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTC 558
              ||||||||   || ||| | |||||| |||| ||||||||||||||||||| || ||||
Db        387 AGATATGTTGTGACAGTCAGTATTGTTATGTGATGCATATTTTTCATTCTGGCATACCTC 446

Qy        559 AAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATG 618
              ||||||| ||   |||| |||| || ||   | || |||| || ||  ||| |||||||
Db        447 AAAGTCTGCAGGTTCTTCTTGAGGACTCCGAAATTATAAAAGTTGGAGTTGCAATTGATG 506

Qy        619 GTGACTCTG 627
              | ||  |||
Db        507 GCGATGCTG 515


RESULT 12
BM358891
LOCUS       BM358891                517 bp    mRNA    linear   EST 09-JAN-2002
DEFINITION  GA__Ea0013O22r Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ea0013O22r, mRNA sequence.
ACCESSION   BM358891
VERSION     BM358891.1  GI:18099637
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
```

```
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total High Quality bases = 327
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence stop: 514.
FEATURES              Location/Qualifiers
     source           1. .517
                      /organism="Gossypium arboreum"
                      /strain="AKA"
                      /cultivar="8400"
                      /db_xref="taxon:29729"
                      /clone="GA__Ea0013O22r"
                      /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                      /tissue_type="Fibers isolated from bolls harvested 7-10
                      dpa"
                      /lab_host="E. coli"
                      /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      135 a    112 c    108 g    161 t      1 others
ORIGIN

  Query Match           16.2%;  Score 169.6;  DB 13;  Length 517;
  Best Local Similarity   60.6%;  Pred. No. 1.1e-31;
  Matches  299;  Conservative    0;  Mismatches  185;  Indels    9;  Gaps    1;

Qy        145 CTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATC 204
              || | ||| || ||||  |   |  |  |  |  |    |  || |    |  |  |
Db         24 CTCCTTCTACTCCTCCTTCTACTCCTTCATCTTCTCGCAAGAAACGCAAGTCCGCCGTCC 83

Qy        205 CAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACAT 264
              |||||     |   || | | | ||||||||||| | ||| ||    ||  |  |
Db         84 ATCATCAAGATAACCTCAAAACCAGCCGCCAATTGCCCCATTCTATTCTCTCACCCCCGC 143

Qy        265 CTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA---------GGAATTTTCCAGCAA 315
              ||| |   |  | | |  || | ||||| | | |||          |  | |||  |
Db        144 CTTCTTTTCCGTCTTCCTTCGCTCGTTGTCAATCTAATACAAGGCTTAGATATCCGCCGT 203

Qy        316 TGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCGAGCAA 375
              |||| ||||| |||   ||||||||||||   || |    |||||| || ||    ||||
Db        204 TGAGATTTGGAGGTCATATTTTGTATAGCTTTACGGAAGATGAGGTAGAAAATGCTGCAA 263

Qy        376 TGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCT 435
              || ||||  | ||| ||  ||| |  ||||   | |||   ||   || |  | | || |
Db        264 TGGAGCTATTAAAAATTGCTGAAATTAAGAAGAAAGAAATGGGTCAAGTTGCTTTAGGGT 323

Qy        436 TGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTG 495
              | ||||||||||||  ||   |||   ||||||  || | || |||||| ||||
Db        324 TTGATATTGAGTGGAAGCCCTCTTTCCAAAAAGGAATTTTGCCTGGGAAGGCTGCGGTGA 383

Qy        496 TCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCC 555
              | |||||||||   || ||| | |||||| |||| ||||||||||||||||||||| || |
Db        384 TGCAGATATGTTGTGACAGTCAGTATTGTTATGTGATGCATATTTTTCATTCTGGCATAC 443

Qy        556 CTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTG 615
              |||| ||||| ||   |||| |||| ||      | || |||| || ||  ||| |||||
Db        444 CTCANAGTCTGCAGGTTCTTCTTGAGGAACTCGAAATTATAAAAGTTGGAGTTGCAATTG 503

Qy        616 ATGGTGACTCTGT 628
              |||| ||  ||||
Db        504 ATGGCGATGCTGT 516


RESULT 13
BM731125
LOCUS       BM731125                555 bp    mRNA    linear   EST 01-MAR-2002
DEFINITION  sal68a05.y1 Gm-c1061 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1061-4233 5' similar to TR:Q9SVM6 Q9SVM6 HYPOTHETICAL 35.1 KD
            PROTEIN. ;, mRNA sequence.
ACCESSION   BM731125
VERSION     BM731125.1  GI:19052458
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact: ccu@resgen.com web site:
            www.resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 424.
FEATURES              Location/Qualifiers
     source           1. .555
                      /organism="Glycine max"
                      /db_xref="taxon:3847"
                      /clone="SOYBEAN CLONE ID: Gm-c1061-4233"
                      /clone_lib="Gm-c1061"
                      /tissue_type="mature flowers of field grown plants"
                      /lab_host="DH10B"
                      /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                      XhoI; The cDNA library was constructed from mRNA isolated
                      from mature flowers of field grown plants for the cultivar
                      Raiden. Complementary DNA was synthesized from mRNA using
                      a primer consisting of a poly(dT) sequence with a XhoI
                      restriction site. EcoRI adapters were ligated to the
                      blunt-ended cDNA fragments followed by XhoI digestion. The
                      cDNA fragments  were directionally cloned into the
                      EcoRI-XhoI restriction site of the pBluescript vector. The
                      ligated cDNA fragments were transformed into DH10B host
                      cells (GibcoBRL). This library was constructed in the
                      laboratory of Dr. Randy Shoemaker."
BASE COUNT      151 a    125 c     97 g    182 t
ORIGIN

  Query Match           16.0%;  Score 167.4;  DB 14;  Length 555;
  Best Local Similarity   60.8%;  Pred. No. 4.1e-31;
  Matches  273;  Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy        231 CGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGT 290
              | | | |   ||  | |  || | | ||  |  || | |    |   ||  | |  |
Db        106 CTCAATTCTGCTATTCACTATACATTTCCTCTAATTTTTCATTTTTCATCTTGCACAGGC 165

Qy        291 TGCCGAGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACT 350
                  |   | ||  || || | ||||| |||| |||||   ||||  ||||||| ||||
Db        166 GATTCAAGAATGACATTGCCTGTAATGAAGTTTAGTGGTCAAATTTCTTATAGCAGGACT 225

Qy        351 GCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGAT 410
              |  ||  || || ||    ||||   ||||  |  || ||||  |    || |
Db        226 TTTGATGCTGTAGAGAAAGCTGCAACAAAGCTCTTACAAATTCTCCAAGAAAAAACGACC 285

Qy        411 GAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGT 470
              ||          | ||    |||| || || ||||||||| ||| |  || |||||||||
Db        286 GACATGATGCAAACTGCAATTGGATTTGACATTGAGTGGAAACCCACCTTCAGAAAAGGT 345

Qy        471 GTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTT 530
              ||||  || || ||||| ||      | ||||||||||  || | ||  ||||| |||||
Db        346 GTTCCTCCCGGAAAGGTAGCAGTGATGCAGATATGTGGTGACACTAGACATTGTCATGTT 405
```

```
Qy        531 ATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACA 590
Db        406 CTACATCTAATTCATTCTGGAATCCCTCAAAATTTACAGCTTTTGCTTGAAGATCCCACA 465

Qy        591 CTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGA 650
Db        466 GTCTTGAAGGTTGGAGCTGGGATTGATGGTGATGCTGTGAAGGTTTTTAGAGATTATAAC 525

Qy        651 GTTAGTATCAAAGATGTTGAGGATCTTTC 679
Db        526 ATATCTGTTAAAGGTGTGACGGATCTTTC 554


RESULT 14
BJ213300
LOCUS       BJ213300                595 bp    mRNA    linear   EST 04-APR-2002
DEFINITION  BJ213300 Y. Ogihara unpublished cDNA library, Wh Triticum aestivum
            cDNA clone wh21m08 5', mRNA sequence.
ACCESSION   BJ213300
VERSION     BJ213300.1  GI:19952807
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 595)
  AUTHORS   Ogihara,Y. and Murai,K.
  TITLE     Expressed genes in Triticum aestivum
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .595
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="wh21m08"
                     /clone_lib="Y. Ogihara unpublished cDNA library, Wh"
                     /tissue_type="spike at meiosis"
                     /dev_stage="Feekes' scale 9"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown under
                     hydroponic conditions at UC Davis, salt stressed  for 12
                     hours, and for 7 days, then dissected and frozen (Akhunov
                     in J  Dvorak Lab). Total RNA was prepared from sheath
                     tissue, equal quantities  of RNA were pooled from the two
                     samples, polyA was purified from the  pooled RNA, a cDNA
                     library was made, and the cDNA clones were in vivo
                     excised to give pBluescript phagemids in the TJ Close lab
                     at the  University of California, Riverside (Akhunov, Chin
                     , Choi, Close, Fenton,  Kianian, Otto, Simons, Zhang).
                     Plasmid DNA preparations and DNA  sequencing were
                     performed in the OD Anderson lab (all other authors)."
BASE COUNT      170 a    119 c    148 g    157 t      1 others
ORIGIN

  Query Match             15.4%;  Score 162;  DB 13;  Length 595;
  Best Local Similarity   58.3%;  Pred. No. 9.1e-30;
  Matches  323;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        321 TTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAG 380
Db         45 TTCAGTGGCAAGATAGTGTACTGCCGGACGGCAATTGAAGTGGAGAAAGCTACGCGGGAG 104

Qy        381 CTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGAT 440
Db        105 ATTGTACGCAAAATCGAAAGCATGA---AGGCCTCTGGCCCGGTCTCCCTTGGTTTCGAT 161

Qy        441 ATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAG 500
Db        162 CTCGAGTGGAAACCCTTTCCCAGAAGAGGAGAACCACCATGTAAAGTCGCGTTGATGCAA 221

Qy        501 ATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAA 560
Db        222 TTATGCATGGACAAAACTCATTGTTATCTCATGCATATCATTCACTCTGGCGTGCCTCCC 281

Qy        561 AGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGT 620
Db        282 ATCTTGAAATCTCTTTTGGAGGACAGTTCATCCGTTAAAGTTGGAGTATGTATAGACAAT 341

Qy        621 GACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCA 680
Db        342 GATGCAAGGAAAATGTTCAATGATTATGATGTCCGTGTACAACCATTGATGGATTTATCA 401

Qy        681 GATTTAGCCAACCAAAAAATTGGTGG---AGATAAAAAATGGGGCCTTGCCTCACTAACT 737
Db        402 ACTGTTGCAAACGTCAAGTTAGCTGGGCCCTATAAAAGATGGAGTCTTGCTGCATTAACT 461

Qy        738 GAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAG 797
Db        462 GAAATGGTTACATGTAAAGAGTTGCCAAAGCCTGGCAACATAAGAATGGGAAACTGGGAG 521

Qy        798 TTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGG 857
Db        522 TCTTTTGTTCTCTCAAAAAAGCAACTTGAGTATGCTGCTACCGATGCCTACATCTCGTGG 581

Qy        858 CATCTTTACAAGGT 871
Db        582 TACTTGTATGAGGT 595


RESULT 15
BI433038
LOCUS       BI433038                499 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  EST535799 P. infestans-challenged leaf Solanum tuberosum cDNA clone
            PPCAZ86 5' sequence, mRNA sequence.
ACCESSION   BI433038
VERSION     BI433038.1  GI:15257728
KEYWORDS    EST.
SOURCE      potato.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Restrepo,S., Griffiths,H.M., Smart,C.D., Cho,J., Chiemingo,A.,
            Bougri,O., Buell,C.R., Ronning,C.M., Fry,W.E. and Baker,B.
  TITLE     Generation of ESTs from Potato Leaves Challenged with Phytophthora
            infestans, Compatible Interaction
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: M13F-R.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="PPCAZ86"
                     /clone_lib="P. infestans-challenged leaf"
                     /tissue_type="leaf"
                     /dev_stage="6 week old"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Whole plants were challenged with 450,000
                     sporangia/ml P. infestans US-1(US 940501) in Biotron
                     (Madison, Wisconsin). Leaf tissue was collected at 1, 2,
```

5, 12, and 24 hours post-challenge and frozen in liquid nitrogen immediately upon removal. Kennebec plants showed no signs of HR. Katahdin plants (susceptible to P. infestans US-1) were used as controls and showed infection. NOTE: We cannot exclude the possibility that this sequence is actually derived from Phytophthora rather than potato."

BASE COUNT 141 a 86 c 118 g 154 t
ORIGIN

Query Match 14.9%; Score 156.8; DB 13; Length 499;
Best Local Similarity 62.5%; Pred. No. 1.8e-28;
Matches 245; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

```
Qy        296 AGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTAC 355
              ||  | | | |  || | ||||  |||  ||    || | || ||||  |||  ||
Db        108 AGTGATGCAATACCCGGAAATGTCTTTTAAAGGACATGTTATATACAGCAGAACTATTAA 167

Qy        356 TGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATC 415
              |||||  || ||    ||| |  ||||  | ||  || ||||    ||||      ||
Db        168 AGAGGTGGAGAAATCTGCAGTAGAGCTGTTAAATTTTGTTGAGGAAAAGAAGAGGAAAGA 227

Qy        416 TGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCT 475
              ||  |  | |  | |||| || || ||||| ||||  || | |||||||| ||||||  |
Db        228 AGGGAATGTTGCTCTTGGATTTGACATTGAATGGAAGCCCACTTTTAGAAGAGGTGTGCC 287

Qy        476 CCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCA 535
              ||  |||||||  ||   | | ||||||||||  || |   | |||||  ||||| ||||
Db        288 ACCTGGGAAGGCCGCTGTTATGCAGATATGTGGTGACAAGGGTAATTGTTATGTTCTGCA 347

Qy        536 TATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGT 595
              |||  | || ||||| ||||||||||| ||| |||  |||| |||| ||| ||||  ||||
Db        348 TATCATCCACTCTGGAATCCCTCAAACTCTGCAATCTCTTCTTGAGGATCCAACTGTTGT 407

Qy        596 AAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAG 655
              |||||  ||| || | ||||    |||  ||   ||  ||  ||| ||  |    ||
Db        408 GAAGGTGGGTGTTTGCATTGCAAATGATGCTTACAAAGTTCGCCAAGATCACAATGTATC 467

Qy        656 TATCAAAGATGTTGAGGATCTTTCAGATTTAG 687
              | | || | | | || |||||||| ||  | |
Db        468 TGTGAAGGCTTTGGAAGATCTTTCTGAACTTG 499
```

Search completed: June 14, 2003, 08:23:16
Job time : 1830 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: June 14, 2003, 07:00:34 ; Search time 68 Seconds
(without alignments)
4730.940 Million cell updates/sec

Title: US-09-896-186B-23
Perfect score: 1049
Sequence: 1 accaaagcattaatttttat.....tttaaccgtccagaaactag 1049

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	111	10.6	5208	3	US-08-781-891-70	Sequence 70, Appl
	2	111	10.6	5208	4	US-09-791-211-11	Sequence 11, Appl
	3	82.8	7.9	6476	4	US-09-127-670-5	Sequence 5, Appli
	4	81.2	7.7	4792	3	US-08-781-891-205	Sequence 205, App
	5	55.4	5.3	98844	4	US-09-791-211-10	Sequence 10, Appl
	6	51.2	4.9	29604	3	US-08-781-891-207	Sequence 207, App
	7	49.2	4.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	8	39.2	3.7	1499	4	US-09-300-672-3	Sequence 3, Appli
	9	39.2	3.7	1500	4	US-09-300-672-1	Sequence 1, Appli
	10	37.2	3.5	5433	3	US-08-929-329-1	Sequence 1, Appli
	11	34.2	3.3	1338	1	US-08-485-216-1	Sequence 1, Appli
	12	34.2	3.3	1338	2	US-09-003-245-1	Sequence 1, Appli
	13	34.2	3.3	1338	3	US-08-853-552-1	Sequence 1, Appli
c	14	34	3.2	2280	4	US-09-346-408-3	Sequence 3, Appli
c	15	33.8	3.2	1272	4	US-09-134-001C-762	Sequence 762, App
	16	33.6	3.2	1011	4	US-09-134-001C-2258	Sequence 2258, Ap
c	17	33	3.1	1444	3	US-09-416-213-1	Sequence 1, Appli
c	18	33	3.1	1444	4	US-09-416-214-1	Sequence 1, Appli
c	19	33	3.1	1444	4	US-09-035-676-1	Sequence 1, Appli
c	20	33	3.1	99500	4	US-09-798-096-10	Sequence 10, Appl
c	21	33	3.1	168575	4	US-09-426-290-1	Sequence 1, Appli
	22	32.6	3.1	5345	1	US-08-044-618-7	Sequence 7, Appli
	23	32.4	3.1	252	2	US-08-623-906A-1	Sequence 1, Appli
	24	32.4	3.1	8703	4	US-08-961-527-177	Sequence 177, App
c	25	32.4	3.1	9636	1	US-08-323-170B-1	Sequence 1, Appli
c	26	32.4	3.1	9636	4	US-08-954-441-1	Sequence 1, Appli
	27	32.4	3.1	14672	4	US-08-961-527-111	Sequence 111, App
	28	32.4	3.1	16442	3	US-08-781-891-208	Sequence 208, App
	29	32.2	3.1	1956	4	US-08-559-896B-1	Sequence 1, Appli
c	30	32	3.1	548	3	US-08-765-332-140	Sequence 140, App
c	31	32	3.1	548	4	US-09-448-894-140	Sequence 140, App
	32	32	3.1	1185	1	US-07-876-280-8	Sequence 8, Appli
	33	32	3.1	1185	1	US-08-049-783-5	Sequence 5, Appli
	34	32	3.1	1185	1	US-08-316-301A-9	Sequence 9, Appli
	35	32	3.1	1185	2	US-08-904-278-9	Sequence 9, Appli
	36	32	3.1	1185	4	US-09-076-137-9	Sequence 9, Appli
	37	32	3.1	1185	4	US-09-222-594-9	Sequence 9, Appli
	38	32	3.1	1185	5	PCT-US92-03624-9	Sequence 9, Appli
c	39	32	3.1	1640	4	US-09-149-476-116	Sequence 116, App
	40	32	3.1	1846	3	US-09-173-581-15	Sequence 15, Appl
	41	32	3.1	1846	4	US-09-420-915-15	Sequence 15, Appl
	42	32	3.1	20303	1	US-08-370-975B-6	Sequence 6, Appli
	43	32	3.1	26764	1	US-08-370-975B-1	Sequence 1, Appli
	44	31.8	3.0	516	3	US-09-023-082A-115	Sequence 115, App
c	45	31.6	3.0	489	1	US-08-334-254-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-781-891-70
; Sequence 70, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5208 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-70

Query Match 10.6%; Score 111; DB 3; Length 5208;
Best Local Similarity 54.1%; Pred. No. 2.5e-25;
Matches 272; Conservative 0; Mismatches 225; Indels 6; Gaps 2;

Qy 396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455

```
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
              |  |  | |||    ||    || || |||||     | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| ||| |  ||||| | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |     |    | |  ||| |||||  |   ||   | |   ||  |  |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||    ||    ||| |   | | || |||      | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932


RESULT 2
US-09-791-211-11
; Sequence 11, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
;  APPLICANT:  Donna T. Ward
;  APPLICANT:  Andrew T. Watt
;  TITLE OF INVENTION:  ANTISENSE MODULATION OF WRN EXPRESSION
;  FILE REFERENCE:  RTS-0205
;  CURRENT APPLICATION NUMBER: US/09/791,211
;  CURRENT FILING DATE:  2001-02-23
;  NUMBER OF SEQ ID NOS:  90
; SEQ ID NO 11
;   LENGTH: 5208
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (232)...(4530)
US-09-791-211-11

  Query Match          10.6%;  Score 111;  DB 4;  Length 5208;
  Best Local Similarity   54.1%;  Pred. No. 2.5e-25;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |     |  ||| |     | | | || || || || || ||||||  ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
              |  |  | |||    ||    || || |||||     | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| ||| |  ||||| | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |     |    | |  ||| |||||  |   ||   | |   ||  |  |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||    ||    ||| |   | | || |||      | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932


RESULT 3
US-09-127-670-5
; Sequence 5, Application US/09127670
; Patent No. 6228583
; GENERAL INFORMATION:
;  APPLICANT: Massachusetts Institute for Technology
;  APPLICANT: Leonard P. Guarente
;  APPLICANT: David A. Sinclair
;  APPLICANT: David B. Lombard
;  TITLE OF INVENTION: ASSAYS FOR COMPOUNDS WHICH EXTEND LIFE
;  TITLE OF INVENTION: SPAN
;  FILE REFERENCE: MIT-7720pA
;  CURRENT APPLICATION NUMBER: US/09/127,670
;  CURRENT FILING DATE: 1998-07-31
;  EARLIER APPLICATION NUMBER: US 60/054,629
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 6476
;   TYPE: DNA
;   ORGANISM: Murine
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (229)...(4432)
US-09-127-670-5

  Query Match          7.9%;  Score 82.8;  DB 4;  Length 6476;
  Best Local Similarity   53.2%;  Pred. No. 3.7e-16;
  Matches  199;  Conservative    0;  Mismatches  172;  Indels    3;  Gaps    1;

Qy        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540
              | |  || ||     |||||  | |||||   |   ||  | |||  |  |  | || ||||
Db        494 GCAGAGTCGCAGTGATCCAGTTGTGTGTGTCTGAGAACAAATGTTACTTGTTTCACATTT 553

Qy        541 TT---CATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAA 597
              |      || | | |||| ||  |  |  ||     |  | ||| |   | || ||   |
Db        554 CTTCCATGTCAGTTTTCCCCCAGGGATTAAAAATGTTACTAGAAAACAAATCAATTAAGA 613

Qy        598 AGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTA 657
              ||| |||  |||| ||||| || |||      ||| ||| | | ||| | ||  || |
Db        614 AGGCAGGGGTTGGGATTGAAGGGGACCAGTGGAAACTTCTGCGTGATTTTGACGTCAAGT 673

Qy        658 TCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAAT 717
              |  |   | ||| ||| ||  | ||| | |||||  ||||  |     | |  | |  |
Db        674 TGGAGAGTTTTGTGGAGCTGACGGATGTTGCCAATGAAAAGTTGAAGTGCGCAGAGACCT 733

Qy        718 GGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAA 777
              || ||||       ||   | |     |  |  | ||| | ||  ||||    ||   |
```

```
Db        734 GGAGCCTCAATGGTCTGGTTAAACACGTCTTAGGGAAACAACTTTTGAAAGACAAGTCCA 793

Qy        778 TCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAA 837
              || |     | || |||     ||  || ||  |  || | |||  ||  || |||||
Db        794 TCCGCTGCAGCAATTGGAGTAATTTCCCCCTCACTGAGGACCAGAAACTGTATGCAGCCA 853

Qy        838 CGGATGCTTATGCT 851
              | ||||||||||||
Db        854 CTGATGCTTATGCT 867


RESULT 4
US-08-781-891-205
; Sequence 205, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  205:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4792 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  145..4347
US-08-781-891-205

  Query Match             7.7%;  Score 81.2;  DB 3;  Length 4792;
  Best Local Similarity   52.3%;  Pred. No. 1e-15;
  Matches  179;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy        510 GATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAA 569
              || || || | ||      ||      |  ||     || | | |||| ||  |  |  ||
Db        442 GAGAGCAAATGTTACTTGTTTCACATTTCTTCCATGTCAGTTTTCCCCCAGGGATTAAAA 501

Qy        570 CATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTG 629
                | | ||| |   | || ||   |||| ||| |||| ||||| || |||      |
Db        502 ATGTTACTAGAAAACAAATCAATTAAGAAGGCAGGGGTTGGGATTGAAGGGGACCAGTGG 561

Qy        630 AAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCC 689
              || ||| | | ||| | ||  || |   |  |   | ||| ||| ||  | ||| | |||
Db        562 AAACTTCTGCGTGATTTTGACGTCAAGTTGGAGAGTTTTGTGGAGCTGACGGATGTTGCC 621

Qy        690 AACCAAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTT 749
              ||  |||| |     | |  | |  ||| ||||       ||   | |     | |
Db        622 AATGAAAAGTTGAAGTGCGCAGAGACCTGGAGCCTCAATGGTCTGGTTAAACACGTCTTA 681

Qy        750 TGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTG 809
              | ||| | ||  ||||    ||    ||| |     | || |||     ||  || ||
Db        682 GGGAAACAACTTTTGAAAGACAAGTCCATCCGCTGCAGCAATTGGAGTAATTTCCCCCTC 741

Qy        810 TCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCT 851
              |  || | |||  ||  || ||||| || ||||||||||||
Db        742 ACTGAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCT 783


RESULT 5
US-09-791-211-10
; Sequence 10, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
;  APPLICANT:  Donna T. Ward
;  APPLICANT:  Andrew T. Watt
;  TITLE OF INVENTION:  ANTISENSE MODULATION OF WRN EXPRESSION
;  FILE REFERENCE:  RTS-0205
;  CURRENT APPLICATION NUMBER: US/09/791,211
;  CURRENT FILING DATE:  2001-02-23
;  NUMBER OF SEQ ID NOS:  90
; SEQ ID NO 10
;  LENGTH: 98844
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: unsure
;  LOCATION: 24962
;  OTHER INFORMATION: unknown
;  NAME/KEY: unsure
;  LOCATION: 64383
;  OTHER INFORMATION: unknown
;  NAME/KEY: unsure
;  LOCATION: 65468
;  OTHER INFORMATION: unknown
;  NAME/KEY: unsure
;  LOCATION: 65469
;  OTHER INFORMATION: unknown
;  NAME/KEY: unsure
;  LOCATION: 65470
;  OTHER INFORMATION: unknown
;  NAME/KEY: unsure
;  LOCATION: 65471
;  OTHER INFORMATION: unknown
;  NAME/KEY: unsure
;  LOCATION: 87130
;  OTHER INFORMATION: unknown
;  NAME/KEY: unsure
;  LOCATION: 89049
;  OTHER INFORMATION: unknown
;  OTHER INFORMATION:
US-09-791-211-10

  Query Match             5.3%;  Score 55.4;  DB 4;  Length 98844;
  Best Local Similarity   60.1%;  Pred. No. 1.4e-06;
  Matches   92;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy        545 TTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGG 604
              | | | | | || ||  |  |  ||    |  ||||| ||  | || ||  ||||| |||
Db      45421 TACAGTTTTTCCCCAGGGATTAAAAATGTTGCTTGAAAATAAAGCAGTTAAAAAGGCAGG 45480

Qy        605 TATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGA 664
              | | |||||||| || ||      ||| ||| | | ||||| ||   | |   | ||  |
Db      45481 TGTAGGAATTGAAGGAGATCAGTGGAAACTTCTACGTGACTTTGATATCAAATTGAAGAA 45540

Qy        665 TGTTGAGGATCTTTCAGATTTAGCCAACCAAAA 697
              | ||| |||  |  ||||| | |||||  ||||
Db      45541 TTTTGTGGAGTTGACAGATGTTGCCAATAAAAA 45573
```

```
RESULT 6
US-08-781-891-207
; Sequence 207, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  207:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  29604 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-891-207

  Query Match             4.9%;  Score 51.2;  DB 3;  Length 29604;
  Best Local Similarity   56.5%;  Pred. No. 1.5e-05;
  Matches   95;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        545 TTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGG 604
              ||  | | |||| ||  |  |  ||    |  | ||| |   | || ||   |||| |||
Db      19789 TTTAGTTTTCCCCCAGGGATTAAAAATGTTACTAGAAAACAAATCAATTAAGAAGGCAGG 19848

Qy        605 TATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGA 664
                |||| ||||| || |||     ||| ||| | | ||| | ||  || |   |  |
Db      19849 GGTTGGGATTGAAGGGGACCAGTGGAAACTTCTGCGTGATTTTGACGTCAAGTTGGAGAG 19908

Qy        665 TGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAA 712
              | ||| ||| ||  | ||| | |||||  ||||  | || | |   ||
Db      19909 TTTTGTGGAGCTGACGGATGTTGCCAATGAAAAGGTAGGCGTAATAAA 19956


RESULT 7
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             4.7%;  Score 49.2;  DB 1;  Length 7218;
  Best Local Similarity   11.5%;  Pred. No. 2.8e-05;
  Matches   33;  Conservative  140;  Mismatches  113;  Indels    0;  Gaps    0;

Qy          9 ATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATCGACGAC 68
              | || ||||  ||  ||||    |     | | |  | ||    |  |||| |  | | |
Db        992 ACTATTTTTCCTTCGTTTGCCATACGCTCACAGAATTAATTCCGAGCTTGGCTGCAGGTC 1051

Qy         69 GCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCC 128
              |      |    ||    :::::: :: ::  : :: ::    :::::: :  :::::::
Db       1052 GAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1111

Qy        129 CGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCAT 188
              : ::::::::::::::::::::: :: :::: :: : :   :: :  ::::: ::: :
Db       1112 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1171

Qy        189 GGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCC 248
              :: :         :::   ::   ::::::  :  : ::: :: ::  :: :::: ::::
Db       1172 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1231

Qy        249 ATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCC 294
              :: ::::::: : :::: :   :  :::::::::::: :: ::
Db       1232 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1277


RESULT 8
US-09-300-672-3
; Sequence 3, Application US/09300672
; Patent No. 6248937
; GENERAL INFORMATION:
;  APPLICANT: Finkelstein, Ruth R.
;  APPLICANT: Lynch, Tim
;  APPLICANT: Goodman, Howard M.
;  APPLICANT: Wang, Ming-Li
```

```
;  TITLE OF INVENTION: A TRANSCRIPTION FACTOR REGULATING SEED DEVELOPMENT,
;  TITLE OF INVENTION: QUALITY AND COLD-TOLERANCE
;  FILE REFERENCE: 480.89(HV)
;  CURRENT APPLICATION NUMBER: US/09/300,672
;  CURRENT FILING DATE: 1999-04-27
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1499
;   TYPE: DNA
;   ORGANISM: Arabidopsis
US-09-300-672-3

  Query Match             3.7%;  Score 39.2;  DB 4;  Length 1499;
  Best Local Similarity   53.2%;  Pred. No. 0.017;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        112 CTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAG 171
              ||  ||  | || |||    |||||||||||||||||| ||| |  ||  |  |     ||
Db        104 CTCTCTCTACTTCCTCTTCCTCTTCTTCTTCTTCTTCTTCATCTATGGACCCTTTAG 163

Qy        172 CTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTC 231
              || |           ||   |||   ||| |||   || |||| ||  |   |||    ||
Db        164 CTTCCCAACATCAACACAACCATCTGGAAGATAATAACCAAACCCTAACCCATAATAATC 223

Qy        232 GCCAATTGCCTCGTTCCATCACTTCTTCTACATCTT 267
                ||||    |    ||| |   || || || || |
Db        224 CTCAATCCGATTCCACCACCGACTCATCAACTTCCT 259


RESULT 9
US-09-300-672-1
; Sequence 1, Application US/09300672
; Patent No. 6248937
; GENERAL INFORMATION:
;  APPLICANT: Finkelstein, Ruth R.
;  APPLICANT: Lynch, Tim
;  APPLICANT: Goodman, Howard M.
;  APPLICANT: Wang, Ming-Li
;  TITLE OF INVENTION: A TRANSCRIPTION FACTOR REGULATING SEED DEVELOPMENT,
;  TITLE OF INVENTION: QUALITY AND COLD-TOLERANCE
;  FILE REFERENCE: 480.89(HV)
;  CURRENT APPLICATION NUMBER: US/09/300,672
;  CURRENT FILING DATE: 1999-04-27
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1500
;   TYPE: DNA
;   ORGANISM: Arabidopsis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (151)..(1134)
US-09-300-672-1

  Query Match             3.7%;  Score 39.2;  DB 4;  Length 1500;
  Best Local Similarity   53.2%;  Pred. No. 0.017;
  Matches   83;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        112 CTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAG 171
              ||  ||  | || |||    |||||||||||||||||| ||| |  ||  |  |     ||
Db        104 CTCTCTCTACTTCCTCTTCCTCTTCTTCTTCTTCTTCTTCATCTATGGACCCTTTAG 163

Qy        172 CTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTC 231
              || |           ||   |||   ||| |||   || |||| ||  |   |||    ||
Db        164 CTTCCCAACATCAACACAACCATCTGGAAGATAATAACCAAACCCTAACCCATAATAATC 223

Qy        232 GCCAATTGCCTCGTTCCATCACTTCTTCTACATCTT 267
                ||||    |    ||| |   || || || || |
Db        224 CTCAATCCGATTCCACCACCGACTCATCAACTTCCT 259


RESULT 10
US-08-929-329-1
; Sequence 1, Application US/08929329
; Patent No. 6120770
;  GENERAL INFORMATION:
;    APPLICANT:  Adams, John H
;    APPLICANT:  Dalton, John P
;    APPLICANT:  Kappe, Stefan
;    TITLE OF INVENTION:  Plasmodium Proteins Useful for Preparing
;    TITLE OF INVENTION:  Vaccine Compositions
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Barnes & Thornburg
;      STREET:  11 S Meridian
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/929,329
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Breen, John P
;      REGISTRATION NUMBER:  38,833
;      REFERENCE/DOCKET NUMBER:  835910-28685
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 231-7745
;      TELEFAX:  (317) 231-7433
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5433 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium yoelii
US-08-929-329-1

  Query Match             3.5%;  Score 37.2;  DB 3;  Length 5433;
  Best Local Similarity   46.0%;  Pred. No. 0.17;
  Matches  126;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        434 CTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGAC 493
              | |  |||||   | |    |   || |  | ||  | |  ||     | |   | |
Db       2912 CATTAATATTCCTTTGTTTTCTTCTTGTTTAGAATATATGTTCATAATGTATCCTTCTGT 2971

Qy        494 TGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTAT 553
              | | |||| |  || | | |  |||||||| | |||   |  |     | | |||| ||
Db       2972 TTTACAGAGAACTGAAAAAAAAAATTATTGGGGTGTGTGGGTTGCAAGTGAATCTGTTAA 3031

Qy        554 CCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAAT 613
                |    | | |  |  ||  ||  | ||| |       | |||| | |     |
Db       3032 TTCGAGTAATTTGTATAATGCTAAAGGAGAATGTTATTATATAAATGAAAAACCTAATTG 3091

Qy        614 TGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGA 673
              || |  ||||   || || | ||||        |    | |  |  || ||| ||||
Db       3092 TGTTATTGACAAAGTAAATCATTTCTCATTTACTTCCCTCACAACAAATGATATTGATTT 3151

Qy        674 TCTTTCAGATTTAGCCAACCAAAAAAATTGGTGGA 707
              |  |  | || |      |  |||| ||| || |
Db       3152 TAATCAAAATATTAATCTCGTAAAACTTGATGAA 3185


RESULT 11
US-08-485-216-1
```

```
; Sequence 1, Application US/08485216
; Patent No. 5695960
;  GENERAL INFORMATION:
;    APPLICANT:  Chan, Voon Loong
;    APPLICANT:  Hani, Eric K.
;    TITLE OF INVENTION:  HIPPURICASE GENE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  3100 No. 5695960west Center, 90 S. 7th Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  U.S.A.
;      ZIP:  55402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,216
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 94/26907
;      FILING DATE:  13-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mueller, Douglas P.
;      REGISTRATION NUMBER:  30,300
;      REFERENCE/DOCKET NUMBER:  7933.29US02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612/332-5300
;      TELEFAX:  612/332-9081
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1338 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1338
US-08-485-216-1

  Query Match             3.3%;  Score 34.2;  DB 1;  Length 1338;
  Best Local Similarity   47.8%;  Pred. No. 0.67;
  Matches   99;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        509 AGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCA 568
              |||| || ||||  || |  | | | |||| |   ||     | |     |  |  | |
Db        798 AGATCCTATTTATGCTGCTTCTTTACTTATTGTAACTTTACAAAGCATAGTATCTTGCAA 857

Qy        569 ACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGT 628
                |  |    || ||||| || ||||||   ||||   |   | || |||  |  || |
Db        858 TGTTGATCCCCAAAATTCAGCAGTTGTAAGCATAGGAGCTTTTAATGCTGGACATGCTTT 917

Qy        629 GAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGC 688
               ||  |  | |  ||   || |   | ||  |    ||||   |   |     ||  | |
Db        918 TAATATCATTCCAGATATTGTAACGATTAAAATGAGTGTTAGAGGATTAGATAATGAAAC 977

Qy        689 CAACCAAAAAATTGGTGGAGATAAAAA 715
                |   |   || ||  | | | |||||
Db        978 TAGAAAGCTAACTGAAGAAAAAAAAAA 1004


RESULT 12
US-09-003-245-1
; Sequence 1, Application US/09003245
; Patent No. 5981189
;  GENERAL INFORMATION:
;    APPLICANT:  Chan, Voon L.
;    APPLICANT:  Hani, Eric K.
;    TITLE OF INVENTION:  Hippuricase Gene
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MERCHANT & GOULD
;      STREET:  3100 No. 5981189west Center, 90 South Seventh Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  U.S.A.
;      ZIP:  55402-4131
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/003,245
;      FILING DATE:  06-JAN-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mueller, Douglas
;      REGISTRATION NUMBER:  30,300
;      REFERENCE/DOCKET NUMBER:  M&G 7933.29-US-D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (612) 332-5300
;      TELEFAX:  (612) 332-9081
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1338 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Campylobacter jejuni
;      STRAIN:  TGH9011
;      INDIVIDUAL ISOLATE:  ATCC43431
;    IMMEDIATE SOURCE:
;      LIBRARY:  C. jejuni genomic library in pBR322
;      CLONE:  pHIP-O
;    POSITION IN GENOME:
;      MAP POSITION:  380 kb SmaI fragment
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1338
US-09-003-245-1

  Query Match             3.3%;  Score 34.2;  DB 2;  Length 1338;
  Best Local Similarity   47.8%;  Pred. No. 0.67;
  Matches   99;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        509 AGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCA 568
              |||| || ||||  || |  | | | |||| |   ||     | |     |  |  | |
Db        798 AGATCCTATTTATGCTGCTTCTTTACTTATTGTAACTTTACAAAGCATAGTATCTTGCAA 857

Qy        569 ACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGT 628
                |  |    || ||||| || ||||||   ||||   |   | || |||  |  || |
Db        858 TGTTGATCCCCAAAATTCAGCAGTTGTAAGCATAGGAGCTTTTAATGCTGGACATGCTTT 917

Qy        629 GAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGC 688
               ||  |  | |  ||   || |   | ||  |    ||||   |   |     ||  | |
Db        918 TAATATCATTCCAGATATTGTAACGATTAAAATGAGTGTTAGAGGATTAGATAATGAAAC 977

Qy        689 CAACCAAAAAATTGGTGGAGATAAAAA 715
                |   |   || ||  | | | |||||
Db        978 TAGAAAGCTAACTGAAGAAAAAAAAAA 1004


RESULT 13
US-08-853-552-1
; Sequence 1, Application US/08853552
; Patent No. 6013501
;  GENERAL INFORMATION:
;    APPLICANT:  Chan, Voon Loong
```

```
;    APPLICANT:  Hani, Eric K.
;    TITLE OF INVENTION:  HIPPURICASE GENE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  3100 No. 6013501west Center, 90 S. 7th Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  U.S.A.
;      ZIP:  55402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/853,552
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/485,216
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mueller, Douglas P.
;      REGISTRATION NUMBER:  30,300
;      REFERENCE/DOCKET NUMBER:  7933.29US02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612/332-5300
;      TELEFAX:  612/332-9081
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1338 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1338
US-08-853-552-1

  Query Match             3.3%;  Score 34.2;  DB 3;  Length 1338;
  Best Local Similarity   47.8%;  Pred. No. 0.67;
  Matches   99;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        509 AGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCA 568
              ||||  || ||||  || |  | | | |||| |   ||     | |     |  |  | |
Db        798 AGATCCTATTTATGCTGCTTCTTTACTTATTGTAACTTTACAAAGCATAGTATCTTGCAA 857

Qy        569 ACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGT 628
                |  |    || ||||| || ||||||   ||||   |   | || |||  |  || |
Db        858 TGTTGATCCCCAAAATTCAGCAGTTGTAAGCATAGGAGCTTTTAATGCTGGACATGCTTT 917

Qy        629 GAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGC 688
               ||  |  | |  ||   || |   | ||  |    ||||   |   |     ||  | |
Db        918 TAATATCATTCCAGATATTGTAACGATTAAAATGAGTGTTAGAGGATTAGATAATGAAAC 977

Qy        689 CAACCAAAAAATTGGTGGAGATAAAAA 715
                |   |   || ||  | | | |||||
Db        978 TAGAAAGCTAACTGAAGAAAAAAAAAA 1004


RESULT 14
US-09-346-408-3/c
; Sequence 3, Application US/09346408B
; Patent No. 6338966
; GENERAL INFORMATION:
;  APPLICANT: Allen, Steve
;  APPLICANT: Anderson, Shawn
;  APPLICANT: Falco, Carl
;  APPLICANT: Rafalski, Antoni
;  TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins
;  FILE REFERENCE: BB-1167-A
;  CURRENT APPLICATION NUMBER: US/09/346,408B
;  CURRENT FILING DATE: 1999-07-01
;  EARLIER APPLICATION NUMBER: 60/092,833
;  EARLIER FILING DATE: July 14, 1998
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
;   LENGTH: 2280
;   TYPE: DNA
;   ORGANISM: Momordica charantia
US-09-346-408-3

  Query Match             3.2%;  Score 34;  DB 4;  Length 2280;
  Best Local Similarity   63.4%;  Pred. No. 1.1;
  Matches   52;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy         10 TTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATCGACGACG 69
              ||  ||||| |||||||  |||   ||||  ||||  | |||||||  |      ||| |
Db       2268 TTTTTTTTTTTTTTTTTTTTTTTTAAATAAAATACCTTCGTCAATTGAATGAGACCATC 2209

Qy         70 CTTTTACAGAGGAAGAGCTTCT 91
              | |||  ||     ||||  ||||
Db       2208 CCTTTTCATCCACAGAGTTTCT 2187


RESULT 15
US-09-134-001C-762/c
; Sequence 762, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 762
;   LENGTH: 1272
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-762

  Query Match             3.2%;  Score 33.8;  DB 4;  Length 1272;
  Best Local Similarity   50.7%;  Pred. No. 0.87;
  Matches  111;  Conservative    0;  Mismatches  102;  Indels    6;  Gaps    1;

Qy        503 ATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAG 562
              | ||  || || || ||  |  ||| ||      |     | || | | || | || | |
Db        318 AAGTCCAGGTAATACTTTATTAGATATTTCTTGTCCACGTAATCCAGTTACCACTAATAT 259

Qy        563 TCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAG------GTATTGGAATTGA 616
              |||||||    |||||  ||   | | |  | |||   |||       ||||| || ||
Db        258 TCTCCAAATATTTATTTGAGCGCCTATATCTATAATAATAGAAATTAATATTGCAAATGC 199

Qy        617 TGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCT 676
                    |    ||   |       ||  || |||| |  ||| | ||| ||      |
Db        198 AAAACTAGCATAAAATTGTGCAGTAAACACTGCAGTTTGAGTCAGAAATGCTGGACCAAT 139

Qy        677 TTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAA 715
                | ||| ||||||   |||||| ||    | |||| ||
Db        138 AGCTGATGTAGCCATTAAAAAAACTGAACCATATAATAA 100


Search completed: June 14, 2003, 08:24:33
Job time : 72 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: June 14, 2003, 07:52:48 ; Search time 219 Seconds
(without alignments)
6936.145 Million cell updates/sec

Title: US-09-896-186B-23
Perfect score: 1049
Sequence: 1 accaaagcattaatttttat.....tttaaccgtccagaaactag 1049

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	111	10.6	5189	9	US-09-292-758-31	Sequence 31, Appl
	2	111	10.6	5189	10	US-09-954-456-1131	Sequence 1131, Ap
	3	44.4	4.2	3149	10	US-09-729-674-127	Sequence 127, App
c	4	41.4	3.9	941	9	US-10-123-155-464	Sequence 464, App
	5	40.8	3.9	13059	7	US-08-781-986A-220	Sequence 220, App
c	6	38.2	3.6	439	10	US-09-864-761-20174	Sequence 20174, A
c	7	37	3.5	1910	9	US-09-938-842A-4570	Sequence 4570, Ap
	8	36.2	3.5	5215	9	US-10-331-061-71	Sequence 71, Appl
	9	36.2	3.5	6177	9	US-10-124-800-3	Sequence 3, Appli
	10	36	3.4	1881	9	US-10-198-846-6503	Sequence 6503, Ap
c	11	36	3.4	99916	10	US-09-816-095-3	Sequence 3, Appli
	12	36	3.4	659158	9	US-09-771-208-20	Sequence 20, Appl
	13	35.8	3.4	29544	10	US-09-464-767-1	Sequence 1, Appli
	14	35.8	3.4	32745	10	US-09-464-767-3	Sequence 3, Appli
	15	35.4	3.4	127197	9	US-09-754-853A-1	Sequence 1, Appli
c	16	35	3.3	954	10	US-09-947-971-3	Sequence 3, Appli
c	17	34.6	3.3	280	9	US-09-933-797-413	Sequence 413, App
c	18	34.4	3.3	463	10	US-09-864-761-5350	Sequence 5350, Ap
	19	34.4	3.3	2000	9	US-09-938-842A-4931	Sequence 4931, Ap
	20	34.4	3.3	53226	10	US-09-818-264-3	Sequence 3, Appli
	21	34	3.2	527	10	US-09-864-761-6599	Sequence 6599, Ap
c	22	34	3.2	527	10	US-09-864-761-8066	Sequence 8066, Ap
c	23	34	3.2	1827	10	US-09-925-302-270	Sequence 270, App
	24	34	3.2	3769	9	US-10-175-523-184	Sequence 184, App
	25	33.8	3.2	556	10	US-09-864-761-8619	Sequence 8619, Ap
	26	33.8	3.2	734	9	US-10-184-644-458	Sequence 458, App
	27	33.8	3.2	734	9	US-10-184-634-458	Sequence 458, App
	28	33.8	3.2	785	10	US-09-770-445-860	Sequence 860, App
	29	33.8	3.2	98829	9	US-10-017-724-3	Sequence 3, Appli
	30	33.6	3.2	258	10	US-09-878-574-10134	Sequence 10134, A
	31	33.6	3.2	285	10	US-09-878-574-7317	Sequence 7317, Ap
	32	33.6	3.2	333	7	US-08-781-986A-4473	Sequence 4473, Ap
c	33	33.6	3.2	5814	9	US-10-092-154-1860	Sequence 1860, Ap
c	34	33.6	3.2	5814	10	US-09-764-847-1860	Sequence 1860, Ap
c	35	33.4	3.2	533	10	US-09-864-761-6759	Sequence 6759, Ap
	36	33.4	3.2	556	9	US-09-918-995-27130	Sequence 27130, A
	37	33.4	3.2	2979	10	US-09-815-242-4754	Sequence 4754, Ap
	38	33.4	3.2	3009	10	US-09-815-242-8801	Sequence 8801, Ap
	39	33.4	3.2	3009	10	US-09-815-242-9060	Sequence 9060, Ap
c	40	33.4	3.2	7035	10	US-09-815-242-8615	Sequence 8615, Ap
	41	33.2	3.2	277	9	US-09-535-459-1301	Sequence 1301, Ap
	42	33.2	3.2	477	10	US-09-864-761-5436	Sequence 5436, Ap
c	43	33.2	3.2	225883	9	US-10-175-523-57	Sequence 57, Appl
	44	33	3.1	305	10	US-09-864-761-19262	Sequence 19262, A
	45	33	3.1	496	10	US-09-864-761-2534	Sequence 2534, Ap

ALIGNMENTS

RESULT 1
US-09-292-758-31
; Sequence 31, Application US/09292758
; Publication No. US20020197602A1
; GENERAL INFORMATION:
; APPLICANT: Burmer, Glenna C.
; APPLICANT: Brown, Joseph P.
; APPLICANT: LifeSpan BioSciences, Inc.
; TITLE OF INVENTION: Nucleic Acid Sequences and Proteins
; TITLE OF INVENTION: Associated With Aging
; FILE REFERENCE: 017473-001110US
; CURRENT APPLICATION NUMBER: US/09/292,758
; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: US 60/081,887
; EARLIER FILING DATE: 1998-04-15
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 5189
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-292-758-31

Query Match 10.6%; Score 111; DB 9; Length 5189;
Best Local Similarity 54.1%; Pred. No. 2e-21;
Matches 272; Conservative 0; Mismatches 225; Indels 6; Gaps 2;

```
Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              |||| | |   |  ||| |    | | | || || || || || |||||| ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  | | |||   ||    || || ||||     | ||| | |||||  |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||     |||
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669
```

```
Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  |||||| | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |     |     | |  ||| ||||| |   ||   | |    ||  | |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||    ||   ||| |   | | || |||      | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932


RESULT 2
US-09-954-456-1131
; Sequence 1131, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1131
;   LENGTH: 5189
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-1131

  Query Match             10.6%;  Score 111;  DB 10;  Length 5189;
  Best Local Similarity   54.1%;  Pred. No. 2e-21;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || |||||| ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
              |  |  | |||   ||   || || |||||    | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  |||||| | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |     |     | |  ||| ||||| |   ||   | |    ||  | |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||    ||   ||| |   | | || |||      | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932


RESULT 3
US-09-729-674-127
; Sequence 127, Application US/09729674
; Patent No. US20010039335A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John M.
;  APPLICANT:  LaVallie, Edward R.
;  APPLICANT:  Collins-Racie, Lisa A.
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Steininger II, Robert J.
;  APPLICANT:  Spaulding, Vikki
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6055-64X
;  CURRENT APPLICATION NUMBER: US/09/729,674
;  CURRENT FILING DATE:  2000-12-04
;  PRIOR APPLICATION NUMBER: 09/539,330
;  PRIOR FILING DATE: 2000-03-30
;  NUMBER OF SEQ ID NOS: 283
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 127
;   LENGTH: 3149
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-729-674-127

  Query Match             4.2%;  Score 44.4;  DB 10;  Length 3149;
  Best Local Similarity   49.3%;  Pred. No. 0.045;
  Matches  148;  Conservative    0;  Mismatches  146;  Indels    6;  Gaps    1;

Qy        571 ATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGA 630
              || || | | ||||   ||  || | || || ||  | |||      |  ||  |   |
Db        470 ATATTTTGGCAGATGGCACCATTTTGAAAGTTGGAGTGGGATGCTCAGAAGATGCCAGCA 529

Qy        631 AGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCA 690
              ||||| | || || |||||  |   | | |  |       ||| ||   | |  ||||||
Db        530 AGCTTCTGCAGGATTATGGCCTCGTTGTTAGGGGGTGCCTGGACCTCCGATACCTAGCCA 589
```

```
Qy        691 ------ACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACAC 744
                  |   ||| | |         |||         ||||    || ||  |||||||
Db        590 TGCGGCAGAGAAACAATTTGCTCTGTAATGGGCTTAGCCTGAAGTCCCTCGCTGAGACTG 649

Qy        745 TTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATC 804
              || |    |        ||    ||| |       | |   |      | ||||||||   | |
Db        650 TTTTGAACTTTCCCCTTGACAAGTCCCTTCTACTTCGTTGCAGCAACTGGGATGCTGAGA 709

Qy        805 CTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTT 864
              ||||  || || | ||| ||    || || || | ||||||  |   ||||  |  ||| |
Db        710 CTCTCACAGAGGACCAGGTAATTTATGCTGCCAGGGATGCCCAGATTTCAGTGGCTCTCT 769


RESULT 4
US-10-123-155-464/c
; Sequence 464, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C30
;  CURRENT APPLICATION NUMBER: US/10/123,155
;  CURRENT FILING DATE:  2002-04-15
; Prior Application removed -  See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 464
;   LENGTH: 941
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-123-155-464

  Query Match             3.9%;  Score 41.4;  DB 9;  Length 941;
  Best Local Similarity   7.7%;  Pred. No. 0.15;
  Matches   58;  Conservative  215;  Mismatches  476;  Indels    0;  Gaps    0;

Qy         68 CGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTC 127
              :::    :   :   :   :|    :  :|          |: :: :   :  :  : :
Db        870 SSSC...M.B.MNWNMY...WT..RCB.NYC..A.......AM.MHCM.S.H..W..M.M 811

Qy        128 CCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCA 187
                 |   :           :   : |  :  |:     |:::     | :  :::
Db        810 N.AYG.T....SM.ASS.S..RMSR..HWC.AS.TCBT.BT.ABHB..S.NCNS.MWMY. 751

Qy        188 TGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTC 247
                : :|  : :: |      :    :::| : :| ::   :  :   |     :  ::
Db        750 RC.TY.BG..RNDBGT.....SY.K..SBSC.HAWA.MH..HY..Y...TM..A.B.NKH 691

Qy        248 CATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATTT 307
              ::   : :       |  : :  | :: :   : |:    :|  :           :   :
Db        690 YM.K.MRK....N.C..B.K..A.DM.R.S.H.TM..S.MC.SB...TNN..STYHNSSB 631

Qy        308 TCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAA 367
              |  | :     : : :   :: ::  : :    : :       :   |:    :::    |
Db        630 TADACM..CA.SHWCHH.RDB.RRCNKCBN.M.W.B......BHAMAM...YDBKNSMSA 571

Qy        368 GCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTT 427
                  ::: :||  | :  : ::  :::: : ::|       :  |:  ::|::::|  :
Db        570 ..A..BDW.RCAH.TCHSCMKRD..MKDBNYCYBA.A....CY..AWANKKAMBHBC..D 511

Qy        428 TGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGT 487
              ::  :::: :     ::  :  |   :  ::: :  :| ::::     : |:   : :
Db        510 WDSSSSSD.SYSG.CHKCMBCHA.G.STKSHW.H.NMANMRSDM..R.B.CSM.THTS.R 451

Qy        488 TGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTC 547
                :  |||::: ::: : :     : | : : :   :: : :  :    : ::
Db        450 .Y.KN..GTCMHRSBHH.K.Y..T.N.B.ASB.DSSN.THB.KTH.GMC..RHCBM...D 391

Qy        548 TGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTAT 607
              ::: :    :     : : :  : :      :   :  |:|       :      |
Db        390 ABSBSB..K.MT.C.N.W.HNWW.KAB.NC.W.DT..DTBABA.C.MSSTSSM.TH...T 331

Qy        608 TGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGT 667
              :|   :  :|    :  |  : :         ::      |:||: |::     : :|
Db        330 S.YRAA.CWN.KTCS..H..TT.H.M...R..S.RH.R....ABTTH.TRHT.N.MH.BT 271

Qy        668 TGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGC 727
              : :   :  : :  :          |  : |:: ::     | |  :   : :  :
Db        270 RBSBMBCSMA.MSBS..HS...TB.RASKMBABH.DH...C.TBABSMB..KNS.T.DY. 211

Qy        728 CTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGG 787
              :: :   :|     :    |:: |     :|  :    :  :|:::  : | :    :
Db        210 .YY.M.S.STM.T..H.G..TKYTTA....AST..Y..C.MAYRASMR.CD.A.S.NCTR 151

Qy        788 GAACTGGGAGTTTTATCCTCTGTCAAAGC 816
              :  : :   :      |  |      :  :
Db        150 D.BBAR..CB.....T..T......D..B 122


RESULT 5
US-08-781-986A-220
; Sequence 220, Application US/08781986A
; Publication No. US20030054436A1
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Staphylococcus aureus Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  5255
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,986A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benson, Bob
;      REGISTRATION NUMBER:  30,446
;      REFERENCE/DOCKET NUMBER:  PB248PP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  220:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13059 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

US-08-781-986A-220

Query Match 3.9%; Score 40.8; DB 7; Length 13059;
Best Local Similarity 53.0%; Pred. No. 1.3;
Matches 87; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy        539 TTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAA 598
              || | |||||  | | | |      |  |     ||||||||||    | ||||  ||
Db      12234 TTCTGATTCTACTGTACTTTGTGAACCACTGATACTTATTGAAGTAGAATCACTGATACT 12293

Qy        599 GGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTAT 658
              |   | | |||  | |||||  ||| | |    |||||  ||||| ||| | |     |
Db      12294 GTCTGATGTTGATAATGATGTCGACACCGATGTGCTTTGTGATGACGATGTACTAGCACT 12353

Qy        659 CAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTG 702
              ||  ||  |||| | |  |   ||| ||   ||  ||  |  ||
Db      12354 CATTGACATTGATGTTGATATCGATGTACTTAAGGAACCAGATG 12397
```

RESULT 6
US-09-864-761-20174/c
; Sequence 20174, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20174
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035419.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-20174

Query Match 3.6%; Score 38.2; DB 10; Length 439;
Best Local Similarity 43.9%; Pred. No. 0.8;
Matches 163; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

```
Qy        340 ATAGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATA 399
              || |  |  ||||| |||| | ||||    | |  |   |   |  |  |    || |
Db        374 ATGGTGATGCTGCTGCTGATGATGATGGTGGGGATAATGATGGTGATGCTGCCGATGGTG 315

Qy        400 CCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTT 459
                  | |||||   |||    |||  || |||  | |||   ||           ||
Db        314 ATGGTGGTGATGATGATGGTGATGCTGCTGATGGTGTTGATGGCGATGCTGCTGATGGTG 255

Qy        460 TTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATT 519
              |      |||| |  |   ||  | | |||  |||          |||    |  | ||
Db        254 ATGATGGTGGTGATTACGATGGTGATGCTGCTGCTGCTGCTGGTGGTGATCCTGCTGATG 195

Qy        520 ATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTG 579
              || ||||| |    ||  |  | || ||| |  |  | |  |    |  | ||  |  ||
Db        194 GTTTTGATGGTTATGATGATGGTGATGCTGCTGACGGTGATGATGGTGATGATGATGATG 135

Qy        580 AAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCC 639
              ||       ||   |||    || ||| | |||||||||    |   |   |
Db        134 GTAATGATGATGATGGTGAGGGTAATAATGGTAATGATGGTGATGGCGATGATGGTGATA 75

Qy        640 ATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAA 699
              || || |||  | |  |    | |||  ||| | |  | |  |   |   |    |   |
Db         74 ATAACGATGATGCTGATGGTGATGATAATGATGGTGATGCTACTGCTGTTGCCACTGATG 15

Qy        700 TTGGTGGAGAT 710
              |||||  |||
Db         14 ATGGTGATGAT 4
```

RESULT 7
US-09-938-842A-4570/c
; Sequence 4570, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4570
; LENGTH: 1910
; TYPE: DNA

```
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-4570

  Query Match             3.5%;  Score 37;  DB 9;  Length 1910;
  Best Local Similarity   53.9%;  Pred. No. 4.7;
  Matches   76;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy          3 CAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATC 62
              |||| ||||   ||  |||||||||||   ||| |||  |       || | | |||  |
Db        653 CAAACCATTGGATTACATTTTTTTGTTCGTGTAGAAGTTAATCAGGCGGCTAGTTGACTT 594

Qy         63 GACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAAT 122
              |||   || |  | |  | || | | || |    |  |   |||||   | |  | | ||
Db        593 GACTTGGCCTGAAAAATGCAAAAACCTCACATATTTAAGCACATCGTGACATGGTCCGAT 534

Qy        123 TTCTCCCGTTCTTCTTCTTCT 143
              || | |  |  ||  | || |
Db        533 TTTTGCTTTCTTTTGTTTTGT 513


RESULT 8
US-10-331-061-71
; Sequence 71, Application US/10331061
; Publication No. US20030101486A1
; GENERAL INFORMATION:
;  APPLICANT: Lassner, Michael
;  APPLICANT:  Metz, James G
;  APPLICANT:  Facciotti, Daniel
;  TITLE OF INVENTION: SCHIZOCHYTRIUM PKS GENES
;  FILE REFERENCE: CGNE.131.02US
;  CURRENT APPLICATION NUMBER: US/10/331,061
;  CURRENT FILING DATE:  2002-12-27
;  PRIOR APPLICATION NUMBER: US/09/231,899
;  PRIOR FILING DATE: 1999-01-14
;  PRIOR APPLICATION NUMBER: 60/048,650
;  PRIOR FILING DATE: 1997-06-04
;  PRIOR APPLICATION NUMBER: 09/090,793
;  PRIOR FILING DATE: 1998-06-04
;  NUMBER OF SEQ ID NOS: 86
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
;   LENGTH: 5215
;   TYPE: DNA
;   ORGANISM: Schizochytrium aggregatum
US-10-331-061-71

  Query Match             3.5%;  Score 36.2;  DB 9;  Length 5215;
  Best Local Similarity   59.0%;  Pred. No. 16;
  Matches   62;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy         83 AGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTC 142
              ||| |   ||   | | | | || ||   | || |    ||  ||   ||||||||||||
Db       3169 AGATCCGATCCTCAACAAGGACAACGCGCCGTCTTCTTCTTCTTCTTCTTCTTCTTC 3228

Qy        143 TTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCA 187
              |||||| ||| || ||||| |         || |  || |||| ||
Db       3229 TTCTTCTTCTTCTTCTCCGTCGCCTGCTCCTTCGGCCCCCGTGCA 3273


RESULT 9
US-10-124-800-3
; Sequence 3, Application US/10124800
; Publication No. US20020194641A1
; GENERAL INFORMATION:
;  APPLICANT: Metz, James
;  APPLICANT:  Barclay, William
;  APPLICANT:  Flatt, James
;  APPLICANT:  Kuner, Jerry
;  TITLE OF INVENTION: PUFA Polyketide Synthase Systems and Uses Thereof
;  FILE REFERENCE: 2997-29
;  CURRENT APPLICATION NUMBER: US/10/124,800
;  CURRENT FILING DATE:  2002-04-16
;  PRIOR APPLICATION NUMBER: 09/231,899
;  PRIOR FILING DATE: 1999-01-14
;  PRIOR APPLICATION NUMBER: 60/284,066
;  PRIOR FILING DATE: 2001-04-16
;  PRIOR APPLICATION NUMBER: 60/298,796
;  PRIOR FILING DATE: 2001-06-15
;  PRIOR APPLICATION NUMBER: 60/323,269
;  PRIOR FILING DATE: 2001-09-18
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 6177
;   TYPE: DNA
;   ORGANISM: Schizochytrium sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(6177)
;   OTHER INFORMATION:
US-10-124-800-3

  Query Match             3.5%;  Score 36.2;  DB 9;  Length 6177;
  Best Local Similarity   59.0%;  Pred. No. 18;
  Matches   62;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy         83 AGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTC 142
              ||| |   ||   | | | | || ||   | || |    ||  ||   ||||||||||||
Db       4479 AGATCCGATCCTCAACAAGGACAACGCGCCGTCTTCTTCTTCTTCTTCTTCTTCTTC 4538

Qy        143 TTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCA 187
              |||||| ||| || ||||| |         || |  || |||| ||
Db       4539 TTCTTCTTCTTCTTCTCCGTCGCCTGCTCCTTCGGCCCCCGTGCA 4583


RESULT 10
US-10-198-846-6503
; Sequence 6503, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6503
;   LENGTH: 1881
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3, 6, 7, 8, 20, 24, 33, 48, 83, 89, 96, 98, 100, 102, 104,
;   LOCATION: 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128,
;   LOCATION: 130, 132, 134, 136, 138, 148, 149, 151, 152, 159, 171, 184,
;   LOCATION: 185, 217, 221, 234, 243, 330, 389, 441, 456, 536, 644
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 701, 743, 884, 921, 931, 993, 1005, 1133, 1142, 1171, 1181,
;   LOCATION: 1202, 1217, 1233, 1260, 1295, 1312, 1316, 1331, 1382, 1386,
;   LOCATION: 1418, 1448, 1459, 1465, 1494, 1501, 1518, 1549, 1588, 1603,
;   LOCATION: 1621, 1635, 1653, 1665, 1670, 1688, 1710, 1721, 1732
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1753, 1774, 1785, 1792, 1801, 1808, 1818, 1823, 1827, 1829,
;   LOCATION: 1834, 1843, 1856, 1867
```

```
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6503

  Query Match             3.4%;  Score 36;  DB 9;  Length 1881;
  Best Local Similarity   48.9%;  Pred. No. 9.2;
  Matches   93;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy        634 TTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACC 693
              ||| | |     | || ||| |  |||| |     |  |   ||   |||| | ||
Db        678 TTTCCAAATTTAAAGGGGTTTTTNCCAAAAAAAAAAAAAAAAAATTTTAATTTCCCAAAAA 737

Qy        694 AAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCA 753
              ||||| | || || || ||||||       ||   || | ||  || | |   | ||  |
Db        738 AAAAANTGGGGGGGGAAAAAAAAAAAATTTTTTTTTCCCAAAGGGAAAAAAAGGGTTTAA 797

Qy        754 AAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAA 813
              || |   |  || ||     |  ||  |   ||   || ||||| |  | | | ||
Db        798 AAAAAAAACCCAAACCCCTTTTTTTTTGGGGTTTAAAAAAGGGGTTTTTTTTTTTTTAAA 857

Qy        814 AGCAGCAGTT 823
              |  |  | ||
Db        858 AAAAAAATTT 867


RESULT 11
US-09-816-095-3/c
; Sequence 3, Application US/09816095
; Patent No. US20020137164A1
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001147
;  CURRENT APPLICATION NUMBER: US/09/816,095
;  CURRENT FILING DATE:  2001-03-26
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 99916
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(99916)
;   OTHER INFORMATION: n = A,T,C or G
US-09-816-095-3

  Query Match             3.4%;  Score 36;  DB 10;  Length 99916;
  Best Local Similarity   46.2%;  Pred. No. 1.3e+02;
  Matches  120;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy        425 TTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAA 484
              || | |||  | | |||  |    | ||     ||  |||      | |   |||  || |
Db      41649 TTCTATTGATTGGAATAGTTTCAGAAGGAATGGTACTAGCTCCTCTTTGTGCCCCTGGTA 41590

Qy        485 GGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCA 544
              |  |  | ||||     | | ||     | |   || ||  | || |  ||   |  ||
Db      41589 GAATTTGGCTGTGAGTCTGTCTGGTCCTGGACTTTTTTTGGTTGGTTGGCTATTAATTAT 41530

Qy        545 TTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGG 604
              | |    ||  |  ||  | ||    | ||  ||  |||||||||   |     | |  |
Db      41529 TGCCTCAATTTCAGAAGCTGTCACTGAACTATTTAGAGATTCAACTTCTTCCTGGTTTAG 41470

Qy        605 TATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGA 664
              | |||| |  |    ||  ||   |||  ||| |  |   | | || ||  |||   |
Db      41469 TCTTGGGAGGGTGTATGTGTCAAGGAATTTTTCCATTTCTTCTAGATTTTCTATTTTATT 41410

Qy        665 TGTTGAGGATCTTTCAGATT 684
              ||   ||     || | | |
Db      41409 TGCGTAGAGGTGTTTATAGT 41390


RESULT 12
US-09-771-208-20
; Sequence 20, Application US/09771208
; Patent No. US20020155564A1
; GENERAL INFORMATION:
;  APPLICANT: MEDRANO, JUAN
;  APPLICANT:  BRADFORD, ERIC
;  APPLICANT:  HORVAT, SIMON
;  TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE
;  FILE REFERENCE: 407T-923710US
;  CURRENT APPLICATION NUMBER: US/09/771,208
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: US 08/999,477
;  PRIOR FILING DATE: 1997-12-29
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
;   LENGTH: 659158
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (123459)..(123478)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (602466)..(602485)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (546998)..(547017)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (494715)..(494814)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (390986)..(391005)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (346860)..(346823)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (317174)..(317193)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (280353)..(280373)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (271829)..(271848)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (183872)..(183891)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (170625)..(170645)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   LOCATION: (132680)..(132700)
;   OTHER INFORMATION: n is unidentified a, c, g, or t
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: n is a, c, g, or t
US-09-771-208-20

  Query Match             3.4%;  Score 36;  DB 9;  Length 659158;
  Best Local Similarity   46.0%;  Pred. No. 4e+02;
  Matches   81;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy        112 CTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAG 171
              | ||||    ||  ||   |||||||||||||||||| ||| || || |  |
Db     390899 CCTCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 390958

Qy        172 CTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTC 231
              || |  ||||   | |                                  |  |  ||  ||
Db     390959 CTTCTTCCTCTTCCTCTTCTTCTTCCTNNNNNNNNNNNNNNNNNNNNNATAACATTCATTC 391018
```

```
Qy        232 GCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCC 287
              ||  | | || |   |   || | ||| |||| | | ||  |  | ||  ||||
Db     391019 TTCATCTACTTCATATTACATCTACCTCTCCATCCTCTTAAACAAATTCTTCCTCC 391074


RESULT 13
US-09-464-767-1
; Sequence 1, Application US/09464767
; Patent No. US20020045249A1
; GENERAL INFORMATION:
;  APPLICANT: Both, Gerald
;  APPLICANT:  Boyle, David
;  APPLICANT:  Vrati, Sudhanshu
;  TITLE OF INVENTION: DNA Encoding Ovine Adenovirus (OAV287) and Its Use as a Viral Vec
;  FILE REFERENCE: 50179-073
;  CURRENT APPLICATION NUMBER: US/09/464,767
;  CURRENT FILING DATE:  1999-12-16
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 29544
;   TYPE: DNA
;   ORGANISM: Ovine adenovirus
US-09-464-767-1

  Query Match             3.4%;  Score 35.8;  DB 10;  Length 29544;
  Best Local Similarity   50.9%;  Pred. No. 65;
  Matches   85;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        442 TTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGA 501
              |||| |  |||||| | |||  | ||   |||  || |       | || ||||
Db       6587 TTGATTCAAGACCATGATTTTCACAAAACGTTGACCAGTATTTTTTAGCTACTGAAATTT 6646

Qy        502 TATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAA 561
              |    | |   |||| ||   | | |||||  ||||    ||||  | |   | | |
Db       6647 GTAGCTCTGTTCTGAATTTTTTAAAAGCTATGCCAATTTCATCTTCTTTTTTATTTAACA 6706

Qy        562 GTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATT 608
                |    |||||  | |       |||  || | ||  ||||| ||||
Db       6707 TTACAAAACATTCTCTGTTTACCTCATAACCTATATCGGTAGCTATT 6753


RESULT 14
US-09-464-767-3
; Sequence 3, Application US/09464767
; Patent No. US20020045249A1
; GENERAL INFORMATION:
;  APPLICANT: Both, Gerald
;  APPLICANT:  Boyle, David
;  APPLICANT:  Vrati, Sudhanshu
;  TITLE OF INVENTION: DNA Encoding Ovine Adenovirus (OAV287) and Its Use as a Viral Vec
;  FILE REFERENCE: 50179-073
;  CURRENT APPLICATION NUMBER: US/09/464,767
;  CURRENT FILING DATE:  1999-12-16
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 32745
;   TYPE: DNA
;   ORGANISM: synthetic construct
US-09-464-767-3

  Query Match             3.4%;  Score 35.8;  DB 10;  Length 32745;
  Best Local Similarity   50.9%;  Pred. No. 70;
  Matches   85;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        442 TTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGA 501
              |||| |  |||||| | |||  | ||   |||  || |       | || ||||
Db       6586 TTGATTCAAGACCATGATTTTCACAAAACGTTGACCAGTATTTTTTAGCTACTGAAATTT 6645

Qy        502 TATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAA 561
              |    | |   |||| ||   | | |||||  ||||    ||||  | |   | | |
Db       6646 GTAGCTCTGTTCTGAATTTTTTAAAAGCTATGCCAATTTCATCTTCTTTTTTATTTAACA 6705

Qy        562 GTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATT 608
                |    |||||  | |       |||  || | ||  ||||| ||||
Db       6706 TTACAAAACATTCTCTGTTTACCTCATAACCTATATCGGTAGCTATT 6752


RESULT 15
US-09-754-853A-1
; Sequence 1, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
;  APPLICANT: Hauge, Brian M.
;  APPLICANT:  Parnell, Laurence D.
;  APPLICANT:  Parsons, Jeremy D.
;  APPLICANT:  Wang, Ming Li
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Soybean Cyst Nematode Resistance
;  FILE REFERENCE: 38-10(15810)B
;  CURRENT APPLICATION NUMBER: US/09/754,853A
;  CURRENT FILING DATE:  2001-01-05
;  PRIOR APPLICATION NUMBER: US 60/174,880
;  PRIOR FILING DATE: 2000-01-07
;  NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1
;   LENGTH: 127197
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 515002_region_G2
US-09-754-853A-1

  Query Match             3.4%;  Score 35.4;  DB 9;  Length 127197;
  Best Local Similarity   50.9%;  Pred. No. 2.2e+02;
  Matches   84;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy        501 ATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAA 560
              |||  |  | | |  || | || |||| ||    ||||||||    |    |||  | ||
Db      18044 ATAGTTAAATAGATAAAATGTTTTGATTTTTGTGATATTTTTTTAACAATAATCGTTGAA 18103

Qy        561 AGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGT 620
              ||   ||| | |||||     || |  ||  | || |    |||   |||  |   ||
Db      18104 TATCAAAAACTTATTATTATTTTTTTAGGACGAGGAAGGTGCTGTAAATGAAAAGGAAGT 18163

Qy        621 GACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGAT 665
              |    ||||| || || |   | |||   |  |  ||  | ||
Db      18164 AATAAATTGAAGTTTATCAAAACCAATGCTTTCGGCTTCTGAAAT 18208


Search completed: June 14, 2003, 09:31:29
Job time : 225 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: June 14, 2003, 07:03:29 ; Search time 2836 Seconds
(without alignments)
9299.883 Million cell updates/sec

Title: US-09-896-186B-23
Perfect score: 1049
Sequence: 1 accaaagcattaatttttat.....tttaaccgtccagaaactag 1049

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*
44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq:*
46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq:*
47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq:*
48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq:*
49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*
50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*
51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*
52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq:*
53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq:*
54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq:*
55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq:*
56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq:*
57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq:*
58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq:*
59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*
60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq:*
61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq:*
62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq:*
63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*
64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq:*
65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq:*
66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq:*
67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq:*
68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq:*
69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq:*
70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq:*
71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq:*
72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq:*
73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq:*
74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq:*
75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq:*
76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq:*
77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq:*
78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq:*
79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq:*
80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq:*
81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq:*
82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq:*
83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq:*
84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq:*
85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1049	100.0	1049	33	US-09-896-186B-23	Sequence 23, Appl
2	1039	99.0	1114	19	US-09-513-996A-67235	Sequence 67235, A
3	1002.2	95.5	1129	19	US-09-513-996A-28855	Sequence 28855, A
4	831	79.2	942	28	US-09-708-427-27137	Sequence 27137, A
5	831	79.2	942	33	US-09-896-186B-1	Sequence 1, Appli
6	471.6	45.0	507	24	US-09-620-393B-6112	Sequence 6112, Ap
c 7	302.2	28.8	118718	20	US-09-534-859-579	Sequence 579, App
c 8	302.2	28.8	118718	31	US-09-803-736-579	Sequence 579, App
9	276.2	26.3	1170	34	US-09-906-226-15	Sequence 15, Appl
10	276.2	26.3	1170	65	US-60-218-993-15	Sequence 15, Appl
11	181.2	17.3	672	23	US-09-605-698-9441	Sequence 9441, Ap
12	174.8	16.7	584	32	US-09-849-529A-2068	Sequence 2068, Ap
13	174.8	16.7	584	63	US-60-196-868-2068	Sequence 2068, Ap
14	174	16.6	1434	28	US-09-708-427-65909	Sequence 65909, A
15	174	16.6	1434	28	US-09-708-427-83271	Sequence 83271, A
c 16	151.2	14.4	130299	31	US-09-803-736-1476	Sequence 1476, Ap
17	142.4	13.6	579	17	US-09-304-517A-210624	Sequence 210624,
18	142.4	13.6	579	17	US-09-371-146A-210624	Sequence 210624,
19	142.4	13.6	579	37	US-09-985-678-210624	Sequence 210624,
20	142.4	13.6	579	56	US-60-125-818-10555	Sequence 10555, A
21	142	13.5	82484	20	US-09-534-859-326	Sequence 326, App

22	142	13.5	82484	31	US-09-803-736-326	Sequence 326, App
23	124.6	11.9	278	16	US-09-263-191-14639	Sequence 14639, A
24	124.6	11.9	278	17	US-09-304-517A-155805	Sequence 155805,
25	124.6	11.9	278	17	US-09-371-146A-155805	Sequence 155805,
26	124.6	11.9	278	36	US-09-975-254-14639	Sequence 14639, A
27	124.6	11.9	278	37	US-09-985-678-155805	Sequence 155805,
28	124.6	11.9	432	25	US-09-654-617-30344	Sequence 30344, A
29	124.6	11.9	432	27	US-09-684-016-30344	Sequence 30344, A
30	113.6	10.8	537	25	US-09-654-617-16008	Sequence 16008, A
31	113.6	10.8	537	27	US-09-684-016-16008	Sequence 16008, A
32	112.2	10.7	697	19	US-09-505-532-9796	Sequence 9796, Ap
33	112.2	10.7	697	31	US-09-819-091A-9796	Sequence 9796, Ap
34	111	10.6	4299	33	US-09-896-186B-17	Sequence 17, Appl
35	111	10.6	5188	71	US-60-278-258-8683	Sequence 8683, Ap
36	111	10.6	5189	1	PCT-US99-08314-31	Sequence 31, Appl
37	111	10.6	5189	16	US-09-292-758-31	Sequence 31, Appl
38	111	10.6	5189	18	US-09-442-589B-1176	Sequence 1176, Ap
39	111	10.6	5189	36	US-09-954-456-1131	Sequence 1131, Ap
40	111	10.6	5189	61	US-60-172-373-7812	Sequence 7812, Ap
41	111	10.6	5189	65	US-60-213-360-4319	Sequence 4319, Ap
42	111	10.6	5189	76	US-60-324-185-27164	Sequence 27164, A
43	111	10.6	5208	1	PCT-US02-03574-11	Sequence 11, Appl
44	111	10.6	5208	9	US-08-594-242-70	Sequence 70, Appl
45	111	10.6	5208	23	US-09-618-166-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1
US-09-896-186B-23
; Sequence 23, Application US/09896186B
; GENERAL INFORMATION:
; APPLICANT: Joshua Z. Levin
; APPLICANT: Ken Phillips
; APPLICANT: Greg Budziszewski
; APPLICANT: Fred Meins
; APPLICANT: Zhenya Glazov
; TITLE OF INVENTION: Methods of Controlling Gene Expression
; FILE REFERENCE: PB/5-31481A
; CURRENT APPLICATION NUMBER: US/09/896,186B
; CURRENT FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 1049
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-896-186B-23

Query Match 100.0%; Score 1049; DB 33; Length 1049;
Best Local Similarity 100.0%; Pred. No. 5.8e-286;
Matches 1049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ACCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGA 60

Qy         61 TCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACA 120

Qy        121 ATTTCTCCCGTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTTCTCCCGTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCT 180

Qy        181 CCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGC 240

Qy        241 CTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA 300

Qy        301 GGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGG 360

Qy        361 TTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAA 420

Qy        421 TAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGG 480

Qy        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540

Qy        541 TTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGG 600

Qy        601 TAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCA 660

Qy        661 AAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGG 720

Qy        721 GCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCA 780

Qy        781 GGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGG 840

Qy        841 ATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTG 900

Qy        901 GCTCATAACGTGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCTCATAACGTGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAA 960

Qy        961 ATGATATGATACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATATGATACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGT 1020

Qy       1021 TCTGGTCCCTTTAACCGTCCAGAAACTAG 1049
              |||||||||||||||||||||||||||||
Db       1021 TCTGGTCCCTTTAACCGTCCAGAAACTAG 1049
```

RESULT 2
US-09-513-996A-67235
; Sequence 67235, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 67235
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..1114
; OTHER INFORMATION: any n or Xaa = unknown

```
;   FEATURE:
;   OTHER INFORMATION: Location 1..1114 / Ceres Seq. ID 2177597
US-09-513-996A-67235

  Query Match             99.0%;  Score 1039;  DB 19;  Length 1114;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-283;
  Matches 1039;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         11 TAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATCGACGACGC 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATCGACGACGC 60

Qy         71 TTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCG 120

Qy        131 TTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGG 180

Qy        191 CCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCAT 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCAT 240

Qy        251 CACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATTTTCC 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATTTTCC 300

Qy        311 AGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCG 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCG 360

Qy        371 AGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGT 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGT 420

Qy        431 TGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGC 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGC 480

Qy        491 GACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGG 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGG 540

Qy        551 TATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGG 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGG 600

Qy        611 AATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGA 660

Qy        671 GGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTC 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTC 720

Qy        731 ACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAA 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAA 780

Qy        791 CTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGC 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGC 840

Qy        851 TTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTGGCTCATAACG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTGGCTCATAACG 900

Qy        911 TGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAAATGATATGAT 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAAATGATATGAT 960

Qy        971 ACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGTTCTGGTCCCT 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGTTCTGGTCCCT 1020

Qy       1031 TTAACCGTCCAGAAACTAG 1049
              ||||||||||||||||||||
Db       1021 TTAACCGTCCAGAAACTAG 1039


RESULT 3
US-09-513-996A-28855
; Sequence 28855, Application US/09513996A
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  ENCODED THEREBY
;  FILE REFERENCE: 2750-709P
;  CURRENT APPLICATION NUMBER: US/09/513,996A
;  CURRENT FILING DATE:  2000-02-25
;  NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 28855
;   LENGTH: 1129
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: 1..1129
;   OTHER INFORMATION: any n'or Xaa = unknown
;   FEATURE:
;   OTHER INFORMATION: Location 1..1129 / Ceres Seq. ID 1572651
US-09-513-996A-28855

  Query Match             95.5%;  Score 1002.2;  DB 19;  Length 1129;
  Best Local Similarity   97.8%;  Pred. No. 1.1e-272;
  Matches 1016;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy         11 TAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATCGACGACGC 70
              | | |||  |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          2 TTAATTTGTTTTTTTTGTTTCATTAAAAGAAATGTCATCGTCAAATTGGATCGACGACGC 61

Qy         71 TTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCT 121

Qy        131 TTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGG 181

Qy        191 CCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCAT 250
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        182 CCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTACCTCGTTCCAT 241

Qy        251 CACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATTTTCC 310
              ||||| | |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        242 CACTTGTCCTACATCTTCTAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATTTTCC 301

Qy        311 AGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCG 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 AGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCG 361

Qy        371 AGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGT 430
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        362 AGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAAGAGCTTTTGT 421

Qy        431 TGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGC 490
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TGGCTTTGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGC 481

Qy        491 GACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGG 550
              |||||||||||||||||||| |||||||||||||||||||||||||||| ||||||||
Db        482 AACTGTCCAGATATGTGTAGATAATAATTATTGTGATGTTATGCATATTTTCCATTCTGG 541

Qy        551 TATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGG 610
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
```

```
Db        542 TATCCCTCAAAGTCTCCAACATCTTATTGAGGATTCAACACTTGTAAAGGTAGGTATTGG 601

Qy        611 AATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGA 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 AATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGA 661

Qy        671 GGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTC 730
              |||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||
Db        662 GGATCTTTCAGATTTAGCCAACCAAAAAATTGATGGAGAGAAAAAATGGGGCCTTGCCTC 721

Qy        731 ACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAA 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 ACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAA 781

Qy        791 CTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGC 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 CTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGC 841

Qy        851 TTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTGGCTCATAACG 910
              |||||||||||||||||||||||||||||||||||||||||||| | |||||||||
Db        842 TTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGCGACTCATAACG 901

Qy        911 TGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAAATGATATGAT 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 TGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAAATGATATGAT 961

Qy        971 ACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGTTCTGGTCCCT 1030
              |||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||
Db        962 ACACCTAATCCAGTCAAGTAGATGCAATTCTTGTGAATATTTTATCTAGTTCTGGTCCCT 1021

Qy       1031 TTAACCGTCCAGAAACTAG 1049
              |||| ||||||||||||||
Db       1022 TTAATCGTCCAGAAACTAG 1040
```

RESULT 4
US-09-708-427-27137
; Sequence 27137, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27137
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..942
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: 1..942
; OTHER INFORMATION: Ceres Seq. ID 1820252
US-09-708-427-27137

Query Match 79.2%; Score 831; DB 28; Length 942;
Best Local Similarity 96.1%; Pred. No. 3.6e-224;
Matches 852; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```
Qy         42 ATGTCATCGTCAAATTGGATCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGAC 101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCATCGTCAAATTGGATCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGAC 60

Qy        102 GCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCCTCTGCTGCTCCG 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCCTCTGCTGCTCCG 120

Qy        162 ACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAAT 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAAT 180

Qy        222 AATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCT 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCT 240

Qy        282 CTCTCCCGTTGCCGAGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTAT 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCTCCCGTTGCCGAGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTAT 300

Qy        342 AGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACC 401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACC 360

Qy        402 AAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTT 461
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTT 420

Qy        462 AGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTAT 521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTAT 480

Qy        522 TGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAA 581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAA 540

Qy        582 GATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCAT 641
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCAT 600

Qy        642 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 701
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 660

Qy        702 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 720

Qy        762 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 780

Qy        822 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGAC 881
              |||||||||||||||||||||||||||||||||||||||||||||||||||    |
Db        781 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTAACAACAACG 840

Qy        882 CTTCCTGATGCTGTCAGTGGCTCATAACGTGAAGGAGGAAGCTTAAA 928
                    ||  | |||    |    | | ||||| |  || || | |
Db        841 AAAAACCATCTTCTCACACTCAACGACCTTGAAGCAAAAATCTCACA 887
```

RESULT 5
US-09-896-186B-1
; Sequence 1, Application US/09896186B
; GENERAL INFORMATION:
; APPLICANT: Joshua Z. Levin
; APPLICANT: Ken Phillips
; APPLICANT: Greg Budziszewski
; APPLICANT: Fred Meins .
; APPLICANT: Zhenya Glazov
; TITLE OF INVENTION: Methods of Controlling Gene Expression
; FILE REFERENCE: PB/5-31481A
; CURRENT APPLICATION NUMBER: US/09/896,186B
; CURRENT FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-896-186B-1

```
  Query Match           79.2%;  Score 831;  DB 33;  Length 942;
  Best Local Similarity  96.1%;  Pred. No. 3.6e-224;
  Matches  852;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         42 ATGTCATCGTCAAATTGGATCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGAC 101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCATCGTCAAATTGGATCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGAC 60

Qy        102 GCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCG 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCG 120

Qy        162 ACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAAT 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAAT 180

Qy        222 AATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCT 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCT 240

Qy        282 CTCTCCCGTTGCCGAGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTAT 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCTCCCGTTGCCGAGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTAT 300

Qy        342 AGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACC 401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACC 360

Qy        402 AAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTT 461
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTT 420

Qy        462 AGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTAT 521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTAT 480

Qy        522 TGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAA 581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAA 540

Qy        582 GATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCAT 641
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCAT 600

Qy        642 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 701
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 660

Qy        702 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 720

Qy        762 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 780

Qy        822 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGAC 881
              ||||||||||||||||||||||||||||||||||||||||||||||||||||    |
Db        781 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTAACAACAACG 840

Qy        882 CTTCCTGATGCTGTCAGTGGCTCATAACGTGAAGGAGGAAGCTTAAA 928
                  ||  | |||    |    | | ||||| |  || || | |
Db        841 AAAAACCATCTTCTCACACTCAACGACCTTGAAGCAAAAATCTCACA 887


RESULT 6
US-09-620-393B-6112
; Sequence 6112, Application US/09620393B
; GENERAL INFORMATION:
;  APPLICANT: ALEXANDROV, Nickolai et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1068P
;  CURRENT APPLICATION NUMBER: US/09/620,393B
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 9948
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6112
;   LENGTH: 507
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..507
;   OTHER INFORMATION: any n = a, g, c, t, unknown, or other
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..507
;   OTHER INFORMATION: Ceres Seq. ID 1392052
US-09-620-393B-6112

  Query Match           45.0%;  Score 471.6;  DB 24;  Length 507;
  Best Local Similarity  97.0%;  Pred. No. 2.1e-122;
  Matches  491;  Conservative    0;  Mismatches   14;  Indels    1;  Gaps    1;

Qy         11 TAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGATCGACGACGC 70
              | | |||  ||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          2 TTAATTTGTTTTTTTTGTTTCATTAAAAGAAATGTCATCGTCAAATTGGATCGACGACGC 61

Qy         71 TTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCG 130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAATTTCTCCCT 121

Qy        131 TTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTCCGTCCATGG 181

Qy        191 CCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCAT 250
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        182 CCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTACCTCGTTCCAT 241

Qy        251 CACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATTTTCC 310
              ||||| | |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        242 CACTTGTCCTACATCTTCTAAACGATTTCCTCTCTCCCGTTGCCGAGCTAGGAATTTTCC 301

Qy        311 AGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCG 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 AGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCG 361

Qy        371 AGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGT 430
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        362 AGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAAGAGCTTTTGT 421

Qy        431 TGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAA-GGTTG 489
              |||||| ||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        422 TGGCTTTGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGGTTG 481

Qy        490 CGACTGTCCAGATATGTGTAGATAGT 515
              | ||||||||||||||||||||||| |
Db        482 CAACTGTCCAGATATGTGTAGATAAT 507


RESULT 7
US-09-534-859-579/c
; Sequence 579, Application US/09534859
; GENERAL INFORMATION:
;  APPLICANT:       Bush, David F.
;  APPLICANT:  Last, Robert L.
;  APPLICANT:  Levin, Irena M.
;  APPLICANT:  Norris, Susan R.
;  APPLICANT:  Parnell, Laurence D.
;  APPLICANT:  Rounsley, Steven D.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION:       PLANT POLYMORPHIC MARKERS AND USES THEREOF
;  FILE REFERENCE:       38-10(15493)B
;  CURRENT APPLICATION NUMBER: US/09/534,859
```

```
;  CURRENT FILING DATE:  2000-03-29
;  NUMBER OF SEQ ID NOS:       1127
; SEQ ID NO 579
;   LENGTH: 118718
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-534-859-579

  Query Match            28.8%;  Score 302.2;  DB 20;  Length 118718;
  Best Local Similarity   97.5%;  Pred. No. 3.5e-73;
  Matches  307;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 CCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGAT 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96598 CCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGAT 96539

Qy         62 CGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAA 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96538 CGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAA 96479

Qy        122 TTTCTCCCGTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTC 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96478 TTTCTCCCGTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTC 96419

Qy        182 CGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCC 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96418 CGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCC 96359

Qy        242 TCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAG 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      96358 TCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGGTAT 96299

Qy        302 GAATTTTCCAGCAAT 316
              |||  |||    |||
Db      96298 GAAACTTCTTCAAAT 96284


RESULT 8
US-09-803-736-579/c
; Sequence 579, Application US/09803736
; GENERAL INFORMATION:
;  APPLICANT: Bush, David F.
;  APPLICANT:  Levin, Irena M.
;  APPLICANT:  Norris, Susan R.
;  APPLICANT:  Rounsley, Steven D.
;  APPLICANT:  Wiegand, Roger C.
;  TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
;  FILE REFERENCE: 38-10(15493)D
;  CURRENT APPLICATION NUMBER: US/09/803,736
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: US 09/534,859
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: identified by Attorney Docket number 04983.0206CPUS01 38-10
;  PRIOR FILING DATE: 2000-10-20
;  NUMBER OF SEQ ID NOS: 1582
; SEQ ID NO 579
;   LENGTH: 118718
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-803-736-579

  Query Match            28.8%;  Score 302.2;  DB 31;  Length 118718;
  Best Local Similarity   97.5%;  Pred. No. 3.5e-73;
  Matches  307;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          2 CCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGAT 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96598 CCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGAT 96539

Qy         62 CGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAA 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96538 CGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACAA 96479

Qy        122 TTTCTCCCGTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTC 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96478 TTTCTCCCGTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCTC 96419

Qy        182 CGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCC 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96418 CGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGCC 96359

Qy        242 TCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTAG 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      96358 TCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGGTAT 96299

Qy        302 GAATTTTCCAGCAAT 316
              |||  |||    |||
Db      96298 GAAACTTCTTCAAAT 96284


RESULT 9
US-09-906-226-15
; Sequence 15, Application US/09906226
; GENERAL INFORMATION:
;  APPLICANT: Butler, Karla
;  APPLICANT:  Cahoon, Rebecca E.
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Sakai, Hajime
;  TITLE OF INVENTION: Plant RNaseD-Like Genes
;  FILE REFERENCE: BB1467 US NA
;  CURRENT APPLICATION NUMBER: US/09/906,226
;  CURRENT FILING DATE:  2001-07-16
;  PRIOR APPLICATION NUMBER: 60/218993
;  PRIOR FILING DATE: July 17, 2000
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
;   LENGTH: 1170
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-906-226-15

  Query Match            26.3%;  Score 276.2;  DB 34;  Length 1170;
  Best Local Similarity   66.6%;  Pred. No. 8.2e-67;
  Matches  395;  Conservative    0;  Mismatches  198;  Indels    0;  Gaps    0;

Qy        300 AGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAG 359
              | ||  || || | ||||| |||| |||||   ||||  ||||||| ||||  |  ||
Db        229 ATGACATTGCCTGTAATGAAGTTTAGTGGTCAAATTTCTTATAGCAGGACTTTTGATGCT 288

Qy        360 GTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGA 419
              || || ||    ||||   ||||  |  || ||||  |    || |     ||
Db        289 GTAGAGAAAGCTGCAACAAAGCTCTTACAAATTCTCCAAGAAAAAACGACCGACATGATG 348

Qy        420 ATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCG 479
              | ||    |||| || || |||||||||| ||| |  || ||||||||||||  ||
Db        349 CAAACTGCAATTGGATTTGACATTGAGTGGAAACCCACCTTCAGAAAAGGTGTTCCTCCC 408

Qy        480 GGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATT 539
              || ||||| ||      | ||||||||||  || | ||   ||||| ||||| | ||| |
Db        409 GGAAAGGTAGCAGTGATGCAGATATGTGGTGACACTAGACATTGTCATGTTCTACATCTA 468

Qy        540 TTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAG 599
              ||||||||| |||||||||| | | || | | |  |||||||| | ||| |  | |||
Db        469 ATTCATTCTGGAATCCCTCAAAATTTACAGCTTTTGCTTGAAGATCCCACAGTCTTGAAG 528

Qy        600 GTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATC 659
              || ||   ||| |||||||||||  |||||||| ||||    || |||    |   | ||
Db        529 GTTGGAGCTGGGATTGATGGTGATGCTGTGAAGGTTTTTAGAGATTATAACATATCTGTT 588

Qy        660 AAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGG 719
              |||| |||   |||||||||   |   || || |||||  |||||||||| | || |||
Db        589 AAAGGTGTGACGGATCTTTCTTTTCATGCTAATCAAAAGCTTGGTGGAGATCATAAGTGG 648

Qy        720 GGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATC 779
              || ||||| ||  | ||||| | |||| | |  ||| ||||    |||||  |||| |||
```

```
Db        649 GGTCTTGCATCTTTGACTGAAAAACTTCTATCAAAACAGCTTAAAAAGCCCAACAAAATA 708

Qy        780 AGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACG 839
              || || || || ||||||  |  |  | |||||||| ||||  || | || || |||||
Db        709 AGACTGGGAAATTGGGAGGCTCCTGTTTTGTCAAAGGAGCAACTAGAGTATGCTGCAACA 768

Qy        840 GATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGC 892
              ||||||| |||||| |||  ||||||  |||   |||| || || || || ||
Db        769 GATGCTTTTGCTTCTTGGTGTCTTTATCAGGCGATTAAAGATCTCCCGGACGC 821


RESULT 10
US-60-218-993-15
; Sequence 15, Application US/60218993
; GENERAL INFORMATION:
;  APPLICANT:  Butler, Karla
;  APPLICANT:  Cahoon, Rebecca E.
;  APPLICANT:  Rafalski, Antoni
;  TITLE OF INVENTION:  Plant RNaseD-Like Genes
;  FILE REFERENCE:  BB1467 US PRV
;  CURRENT APPLICATION NUMBER: US/60/218,993
;  CURRENT FILING DATE:  2000-07-17
;  NUMBER OF SEQ ID NOS:  22
;  SOFTWARE:  Microsoft Office 97
; SEQ ID NO 15
;   LENGTH: 1170
;   TYPE: DNA
;   ORGANISM: Glycine max
US-60-218-993-15

  Query Match             26.3%;  Score 276.2;  DB 65;  Length 1170;
  Best Local Similarity   66.6%;  Pred. No. 8.2e-67;
  Matches  395;  Conservative    0;  Mismatches  198;  Indels    0;  Gaps    0;

Qy        300 AGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAG 359
              | ||  || || | ||||| |||| ||||   ||||  ||||||| ||||  |  ||
Db        229 ATGACATTGCCTGTAATGAAGTTTAGTGGTCAAATTTCTTATAGCAGGACTTTTGATGCT 288

Qy        360 GTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGA 419
              || || ||    ||||   ||||  |  || ||||  |    || |     ||
Db        289 GTAGAGAAAGCTGCAACAAAGCTCTTACAAATTCTCCAAGAAAAAACGACCGACATGATG 348

Qy        420 ATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCG 479
                | ||    |||| || || |||||||||| ||| |  || ||||||||||||  ||
Db        349 CAAACTGCAATTGGATTTGACATTGAGTGGAAACCCACCTTCAGAAAAGGTGTTCCTCCC 408

Qy        480 GGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATT 539
              || ||||| ||     | ||||||||||  || | ||   |||| |||| | ||| |
Db        409 GGAAAGGTAGCAGTGATGCAGATATGTGGTGACACTAGACATTGTCATGTTCTACATCTA 468

Qy        540 TTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAG 599
               |||||||||| |||||||||| | | || | | |  |||||||| | ||| |  | |||
Db        469 ATTCATTCTGGAATCCCTCAAAATTTACAGCTTTTGCTTGAAGATCCCACAGTCTTGAAG 528

Qy        600 GTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATC 659
              || ||   ||| |||||||||||  |||||||| |||    || |||    |  | |
Db        529 GTTGGAGCTGGGATTGATGGTGATGCTGTGAAGGTTTTTAGAGATTATAACATATCTGTT 588

Qy        660 AAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGG 719
              |||| |||   |||||||||    |   || || |||||  |||||||||| | || |||
Db        589 AAAGGTGTGACGGATCTTTCTTTTCATGCTAATCAAAAGCTTGGTGGAGATCATAAGTGG 648

Qy        720 GGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATC 779
              || ||||| ||  | ||||| | |||| | |  ||| ||||    |||||  |||| |||
Db        649 GGTCTTGCATCTTTGACTGAAAAACTTCTATCAAAACAGCTTAAAAAGCCCAACAAAATA 708

Qy        780 AGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACG 839
              || || || || ||||||  |  |  | |||||||| ||||  || | || || |||||
Db        709 AGACTGGGAAATTGGGAGGCTCCTGTTTTGTCAAAGGAGCAACTAGAGTATGCTGCAACA 768

Qy        840 GATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGC 892
              ||||||| |||||| |||  ||||||  |||   |||| || || || || ||
Db        769 GATGCTTTTGCTTCTTGGTGTCTTTATCAGGCGATTAAAGATCTCCCGGACGC 821


RESULT 11
US-09-605-698-9441
; Sequence 9441, Application US/09605698
; GENERAL INFORMATION:
;  APPLICANT:  Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael A.
;  TITLE OF INVENTION:  Polynucleotides, Materials Incorporating
;  TITLE OF INVENTION:  Them, and Methods for Using Them
;  FILE REFERENCE:  11000.1041U1
;  CURRENT APPLICATION NUMBER: US/09/605,698
;  CURRENT FILING DATE:  2000-06-21
;  NUMBER OF SEQ ID NOS:  25120
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 9441
;   LENGTH: 672
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-605-698-9441

  Query Match             17.3%;  Score 181.2;  DB 23;  Length 672;
  Best Local Similarity   60.9%;  Pred. No. 5.4e-40;
  Matches  313;  Conservative    0;  Mismatches  198;  Indels    3;  Gaps    1;

Qy        382 TTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATA 441
              || | ||     |  || ||| || |  || |   |    ||   |  ||||||| || |
Db          1 TTTTGAAGAAAATCAATGCCATGAAAAGTGTAATGGATCAAGTCATAATTGGCTTTGACA 60

Qy        442 TTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGA 501
              ||||||  || || |  |||  |||||||||||  ||  |||||   ||     | ||||
Db         61 TTGAGTATAGGCCCACATTTCAAAAAGGTGTTCAACCTCGGAAGACGGCATTAATGCAGA 120

Qy        502 TATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAA 561
              | ||||  || | | |  | ||  |||| |||||| || ||||||||||  | ||||  |
Db        121 TTTGTGGTGACAATGACCACTGCCATGTGATGCATGTTATTCATTCTGGCGTACCTCCCA 180

Qy        562 GTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTG 621
              || |  | | | || | || ||| ||||  |  |||||||  ||    || || | |||||
Db        181 GTTTGAAGCTTATTTTAGAGGATACAACTGTATTAAAGGTTGGAGCAGGCATAGCTGGTG 240

Qy        622 ACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAG 681
              |  ||  |||| |  |  | || ||    ||   | |||| | | || ||||||||||||
Db        241 ATGCTAGGAAGGTGCTTAAGGATTACAATGTATCTGTCAATGCTTTTCAGGATCTTTCAG 300

Qy        682 ATTTAGCCAACCAAAAAAATTG---GTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTG 738
              |  | || || ||||  |||           | | |  |||| |||| || ||||| |
Db        301 AACTTGCTAATCAAAAGCTTGCCATGCCTCGTTATAGGTGGGGTCTTGCATCTCTAACCG 360

Qy        739 AGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGT 798
              ||   ||| |    ||  ||||   ||    |  | || |||   || || ||||||
Db        361 AGGTGCTTCTAAATAAGCAGCTGAACAAATTCATGAAGATTAGGTGCGGAAATTGGGAGG 420

Qy        799 TTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGC 858
               |  | | | |||||  | ||   ||  || || || || || || || | ||||||||||
Db        421 CTCTTTTTTTATCAAAAGAACAACAACTTTATGCTGCCACTGACGCATTTGCTTCTTGGC 480

Qy        859 ATCTTTACAAGGTTCTTAAGGACCTTCCTGATGC 892
              |||| ||  ||||  | |||  ||| || |||||
Db        481 ATCTATATCAGGTCTTAAAGAGCCTGCCAGATGC 514


RESULT 12
US-09-849-529A-2068
; Sequence 2068, Application US/09849529A
; GENERAL INFORMATION:
;  APPLICANT: Fincher, Karen L.
;  APPLICANT:  Ziegler, Todd E.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(51893)B
;  CURRENT APPLICATION NUMBER: US/09/849,529A
;  CURRENT FILING DATE:  2001-05-07
```

```
;  PRIOR APPLICATION NUMBER: US 60/196,868
;  PRIOR FILING DATE: 2000-05-09
;  NUMBER OF SEQ ID NOS: 24076
; SEQ ID NO 2068
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(584)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: LIB3543-036-Q1-K1-F5
US-09-849-529A-2068

  Query Match           16.7%;  Score 174.8;  DB 32;  Length 584;
  Best Local Similarity   64.4%;  Pred. No. 3.3e-38;
  Matches  282;  Conservative    0;  Mismatches  147;  Indels    9;  Gaps    1;

Qy        208 ATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTT 267
              |||||     |   || | | | |||||||||||| | ||| ||    ||  |  |  |||
Db        147 ATCAAGATAACCTCAAAACCAGCCGCCAATTGCCCCATTCTATTCTCTCACCCCCGCCTT 206

Qy        268 ATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA---------GGAATTTTCCAGCAATGA 318
              |   |  | | |  || | ||||| | | |||          |  | |||  |  |||
Db        207 CTTTTCCGTCTTCCTTCGCTCGTTGTCAATCTAATACAAGGCTTAGATATCCGCCGTTGA 266

Qy        319 GGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGC 378
              | ||||| |||   |||||||||||   || |    ||||| || ||    ||||||
Db        267 GATTTGGAGGTCATATTTTGTATAGCTTTACGGAAGATGAGGTAGAAAATGCTGCAATGG 326

Qy        379 AGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGG 438
              ||||  | ||| || |||| |  ||||   | |||   ||   || |  | | || || |
Db        327 AGCTATTAAAAATTGTTGAAATTAAGAAGAAAGAAATGGGTCAAGTTGCTTTAGGGTTTG 386

Qy        439 ATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCC 498
              ||||||||||||  ||   |||   |||||  || | || ||||||| ||||    | |
Db        387 ATATTGAGTGGAAGCCCTCTTTCCAAAAAGGAATTTTGCCTGGGAAGGCTGCGGTGATGC 446

Qy        499 AGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTC 558
              ||||||||   || ||| | |||||| |||| |||||||||||||||||||| || ||||
Db        447 AGATATGTTGTGACAGTCAGTATTGTTATGTGATGCATATTTTTCATTCTGGCATACCTC 506

Qy        559 AAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATG 618
              ||||||| ||   |||| |||| || ||   | || |||| || ||  ||| ||||||||
Db        507 AAAGTCTGCAGGTTCTTCTTGAGGACTCCGAAATTATAAAAGTTGGAGTTGCAATTGATG 566

Qy        619 GTGACTCTGTGAAGCTTT 636
              | ||  || | ||| | |
Db        567 GCGATGCTATCAAGGTGT 584


RESULT 13
US-60-196-868-2068
; Sequence 2068, Application US/60196868
; GENERAL INFORMATION:
;  APPLICANT:       Fincher, Karen L.
;  APPLICANT:  McCarter, David W.
;  APPLICANT:  Ziegler, Todd E.
;  TITLE OF INVENTION:       NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE:       38-21(51893)A
;  CURRENT APPLICATION NUMBER: US/60/196,868
;  CURRENT FILING DATE:  2000-05-09
;  NUMBER OF SEQ ID NOS:       24054
; SEQ ID NO 2068
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   OTHER INFORMATION: Clone ID: LIB3543-036-Q1-K1-F5
US-60-196-868-2068

  Query Match           16.7%;  Score 174.8;  DB 63;  Length 584;
  Best Local Similarity   64.4%;  Pred. No. 3.3e-38;
  Matches  282;  Conservative    0;  Mismatches  147;  Indels    9;  Gaps    1;

Qy        208 ATCAAATCCCCAATAATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTT 267
              |||||     |   || | | | |||||||||||| | ||| ||    ||  |  |  |||
Db        147 ATCAAGATAACCTCAAAACCAGCCGCCAATTGCCCCATTCTATTCTCTCACCCCCGCCTT 206

Qy        268 ATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA---------GGAATTTTCCAGCAATGA 318
              |   |  | | |  || | ||||| | | |||          |  | |||  |  |||
Db        207 CTTTTCCGTCTTCCTTCGCTCGTTGTCAATCTAATACAAGGCTTAGATATCCGCCGTTGA 266

Qy        319 GGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGC 378
              | ||||| |||   |||||||||||   || |    ||||| || ||    ||||||
Db        267 GATTTGGAGGTCATATTTTGTATAGCTTTACGGAAGATGAGGTAGAAAATGCTGCAATGG 326

Qy        379 AGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGG 438
              ||||  | ||| || |||| |  ||||   | |||   ||   || |  | | || || |
Db        327 AGCTATTAAAAATTGTTGAAATTAAGAAGAAAGAAATGGGTCAAGTTGCTTTAGGGTTTG 386

Qy        439 ATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCC 498
              ||||||||||||  ||   |||   |||||  || | || ||||||| ||||    | |
Db        387 ATATTGAGTGGAAGCCCTCTTTCCAAAAAGGAATTTTGCCTGGGAAGGCTGCGGTGATGC 446

Qy        499 AGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTC 558
              ||||||||   || ||| | |||||| |||| |||||||||||||||||||| || ||||
Db        447 AGATATGTTGTGACAGTCAGTATTGTTATGTGATGCATATTTTTCATTCTGGCATACCTC 506

Qy        559 AAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATG 618
              ||||||| ||   |||| |||| || ||   | || |||| || ||  ||| ||||||||
Db        507 AAAGTCTGCAGGTTCTTCTTGAGGACTCCGAAATTATAAAAGTTGGAGTTGCAATTGATG 566

Qy        619 GTGACTCTGTGAAGCTTT 636
              | ||  || | ||| | |
Db        567 GCGATGCTATCAAGGTGT 584


RESULT 14
US-09-708-427-65909
; Sequence 65909, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65909
;   LENGTH: 1434
;   TYPE: DNA
;   ORGANISM: Zea mays subsp. mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..1434
;   OTHER INFORMATION: any n = a, g, c, t, unknown, or other
;   NAME/KEY: misc_feature
;   LOCATION: 1..1434
;   OTHER INFORMATION: Ceres Seq. ID 1929706
US-09-708-427-65909

  Query Match           16.6%;  Score 174;  DB 28;  Length 1434;
  Best Local Similarity   57.4%;  Pred. No. 8.6e-38;
  Matches  334;  Conservative    0;  Mismatches  245;  Indels    3;  Gaps    1;

Qy        309 CCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAG 368
              | |  |||    ||  | |||| ||| | ||  ||| |||  |||| ||||  || ||
Db        376 CAACAAATATCATTCAGCGGTAAGATAGTTTACTGCAGGACACCTACAGAGGCGGAGAAA 435

Qy        369 CGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTT 428
                ||||   | | |         |||| |  | || |       | ||    | || |
Db        436 GCTGCAACTGACATCTTACTCAAAATTGAGAGGATGAAA---ACCCCAGGCCAGGTTTCT 492

Qy        429 GTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTT 488
```

```
Db        493 CTTGGATTTGATCTCGAGTGGAGGCCCTTTCCAAGAAGAGGAGAACCACCTTGTAAGGTT 552

Qy        489 GCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCT 548

Db        553 GCTGTAATGCAACTATGCATGGAGAAAACTCTGTGCTATGTCCTGCATATTGCTCACTCT 612

Qy        549 GGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATT 608

Db        613 GGGGTACCGCCTATACTGAAAACTCTTTTGGAGGATAGTTCGTCCATTAAAGTTGGAATA 672

Qy        609 GGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTT 668

Db        673 TGCATAGACAATGATGCAAGGAAAATGTTTAATGACTATGATGTCTGTGTACAACCACTA 732

Qy        669 GAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCC 728

Db        733 ATGGATTTGTCGACCTTGGCAAATGTCAAGTTAGCTTCTCCCAAAAGATGGAGCCTTGCT 792

Qy        729 TCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGG 788

Db        793 TCTTTAACCGAAATGATCACATGCAAAGAGTTGCCAAAGCCGAGCAACATAAGAATGGGG 852

Qy        789 AACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTAT 848

Db        853 AACTGGGAGGCTGATGTCCTCTCCAAACAACAACTTCAATATGCTGCTACAGATGCTTAT 912

Qy        849 GCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGAT 890

Db        913 ATCTCATGGTACTTGTATGAGGCCCTGCAGACTCTTCCAGAT 954
```

RESULT 15
US-09-708-427-83271
; Sequence 83271, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 83271
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1434
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: 1..1434
; OTHER INFORMATION: Ceres Seq. ID 1962403
US-09-708-427-83271

Query Match 16.6%; Score 174; DB 28; Length 1434;
Best Local Similarity 57.4%; Pred. No. 8.6e-38;
Matches 334; Conservative 0; Mismatches 245; Indels 3; Gaps 1;

```
Qy        309 CCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGGTTGATAAG 368

Db        376 CAACAAATATCATTCAGCGGTAAGATAGTTTACTGCAGGACACCTACAGAGGCGGAGAAA 435

Qy        369 CGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTT 428

Db        436 GCTGCAACTGACATCTTACTCAAAATTGAGAGGATGAAA---ACCCCAGGCCAGGTTTCT 492

Qy        429 GTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTT 488

Db        493 CTTGGATTTGATCTCGAGTGGAGGCCCTTTCCAAGAAGAGGAGAACCACCTTGTAAGGTT 552

Qy        489 GCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCT 548

Db        553 GCTGTAATGCAACTATGCATGGAGAAAACTCTGTGCTATGTCCTGCATATTGCTCACTCT 612

Qy        549 GGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATT 608

Db        613 GGGGTACCGCCTATACTGAAAACTCTTTTGGAGGATAGTTCGTCCATTAAAGTTGGAATA 672

Qy        609 GGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTT 668

Db        673 TGCATAGACAATGATGCAAGGAAAATGTTTAATGACTATGATGTCTGTGTACAACCACTA 732

Qy        669 GAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCC 728

Db        733 ATGGATTTGTCGACCTTGGCAAATGTCAAGTTAGCTTCTCCCAAAAGATGGAGCCTTGCT 792

Qy        729 TCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGG 788

Db        793 TCTTTAACCGAAATGATCACATGCAAAGAGTTGCCAAAGCCGAGCAACATAAGAATGGGG 852

Qy        789 AACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTAT 848

Db        853 AACTGGGAGGCTGATGTCCTCTCCAAACAACAACTTCAATATGCTGCTACAGATGCTTAT 912

Qy        849 GCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGAT 890

Db        913 ATCTCATGGTACTTGTATGAGGCCCTGCAGACTCTTCCAGAT 954
```

Search completed: June 14, 2003, 09:12:07
Job time : 2843 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: June 14, 2003, 07:04:53 ; Search time 914 Seconds
(without alignments)
7928.765 Million cell updates/sec

Title: US-09-896-186B-23
Perfect score: 1049
Sequence: 1 accaaagcattaatttttat.....tttaaccgtccagaaactag 1049

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7791535 seqs, 3454190175 residues

Total number of hits satisfying chosen parameters: 15583070

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq2:*
7: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq3:*
8: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq2:*
10: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1049	100.0	1049	5	US-09-896-186C-23	Sequence 23, Appl
2	831	79.2	942	5	US-09-896-186C-1	Sequence 1, Appli
3	276.2	26.3	1253	8	US-10-424-599-84537	Sequence 84537, A
4	181.2	17.3	672	8	US-10-389-048-9441	Sequence 9441, Ap
5	142.4	13.6	579	5	US-09-531-113-12947	Sequence 12947, A
6	142.4	13.6	579	6	US-09-531-113-12947	Sequence 12947, A
7	124.6	11.9	278	5	US-09-531-113-705	Sequence 705, App
8	124.6	11.9	278	6	US-09-531-113-705	Sequence 705, App
9	124.6	11.9	278	6	US-09-615-606A-75994	Sequence 75994, A
10	111	10.6	3255	6	US-09-724-676-40395	Sequence 40395, A
11	111	10.6	3255	6	US-09-724-676A-40395	Sequence 40395, A
12	111	10.6	4299	5	US-09-896-186C-17	Sequence 17, Appl
13	111	10.6	5208	6	US-09-949-001-3	Sequence 3, Appli
14	111	10.6	5208	8	US-10-374-077-70	Sequence 70, Appl
15	111	10.6	5365	6	US-09-949-001-9	Sequence 9, Appli
16	111	10.6	5365	8	US-10-170-235-31805	Sequence 31805, A
17	105.8	10.1	824	8	US-10-437-963-15803	Sequence 15803, A
18	95.6	9.1	724	8	US-10-424-599-86601	Sequence 86601, A
19	82.8	7.9	4939	9	US-10-144-771-243	Sequence 243, App
20	81.2	7.7	4792	8	US-10-374-077-205	Sequence 205, App
21	74.2	7.1	271	5	US-09-531-113-22352	Sequence 22352, A
22	74.2	7.1	271	6	US-09-531-113-22352	Sequence 22352, A
23	74.2	7.1	271	6	US-09-615-606A-86495	Sequence 86495, A
24	56.4	5.4	510	5	US-09-837-604B-53366	Sequence 53366, A
25	56.4	5.4	510	7	US-09-837-604A-53366	Sequence 53366, A
26	55.4	5.3	143776	6	US-09-949-001-29	Sequence 29, Appl
27	55.4	5.3	144034	6	US-09-949-001-35	Sequence 35, Appl
28	51.2	4.9	610	8	US-10-424-599-76746	Sequence 76746, A
29	51.2	4.9	29604	8	US-10-374-077-207	Sequence 207, App
30	46	4.4	416	8	US-10-205-189A-3109	Sequence 3109, Ap
31	45.2	4.3	601	6	US-09-949-001-104	Sequence 104, App
32	45.2	4.3	601	6	US-09-949-001-560	Sequence 560, App
33	44.4	4.2	449	9	US-10-276-774-284	Sequence 284, App
34	44.4	4.2	4461	8	US-10-170-235-21239	Sequence 21239, A
35	44.4	4.2	4950	8	US-10-170-235-21355	Sequence 21355, A
36	42.2	4.0	1141	6	US-09-806-708B-22	Sequence 22, Appl
37	41.6	4.0	247475	10	US-60-465-241-52767	Sequence 52767, A
c 38	41.4	3.9	18688	5	US-09-948-128-171	Sequence 171, App
39	40.8	3.9	852	6	US-09-950-084-3306	Sequence 3306, Ap
c 40	40.8	3.9	5991	6	US-09-950-084-2617	Sequence 2617, Ap
c 41	40.8	3.9	6816	9	US-10-282-122A-7740	Sequence 7740, Ap
c 42	40.8	3.9	6852	9	US-10-172-502-3	Sequence 3, Appli
43	40.8	3.9	13059	9	US-10-329-624-220	Sequence 220, App
44	40.8	3.9	812842	6	US-09-950-084-7449	Sequence 7449, Ap
45	40.4	3.9	15674	9	US-10-240-485-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-09-896-186C-23
; Sequence 23, Application US/09896186C
; GENERAL INFORMATION:
; APPLICANT: Joshua Z. Levin
; APPLICANT: Ken Phillips
; APPLICANT: Greg Budziszewski
; APPLICANT: Fred Meins
; APPLICANT: Zhenya Glazov
; TITLE OF INVENTION: Methods of Controlling Gene Expression
; FILE REFERENCE: 31481USNP
; CURRENT APPLICATION NUMBER: US/09/896,186C
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/222,202
; PRIOR FILING DATE: 2000-08-01
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 1049
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-896-186C-23

Query Match 100.0%; Score 1049; DB 5; Length 1049;
Best Local Similarity 100.0%; Pred. No. 1e-279;
Matches 1049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ACCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCAAAGCATTAATTTTTATTTTTTTGTTTCAGTAAAAGAAATGTCATCGTCAAATTGGA 60

Qy         61 TCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGACGCCATCGAAGCTTCCTACA 120

Qy        121 ATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCGACCGTACAAGCTACAACCT 180

Qy        181 CCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAATAATATCCGTCGCCAATTGC 240

Qy        241 CTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 CTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCTCTCTCCCGTTGCCGAGCTA 300

Qy        301 GGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAGG 360

Qy        361 TTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAA 420

Qy        421 TAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGG 480

Qy        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATTT 540

Qy        541 TTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGG 600

Qy        601 TAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCA 660

Qy        661 AAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGGG 720

Qy        721 GCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCA 780

Qy        781 GGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGG 840

Qy        841 ATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGCTGTCAGTG 900

Qy        901 GCTCATAACGTGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCTCATAACGTGAAGGAGGAAGCTTAAAGGTTAGCCTATAACCCCAAGAGTTAGCATCAA 960

Qy        961 ATGATATGATACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGATATGATACACCTAATCTAGTCAAGTAGATGCAATTCTTGTGAATATTGTATCTAGT 1020

Qy       1021 TCTGGTCCCTTTAACCGTCCAGAAACTAG 1049
              |||||||||||||||||||||||||||||
Db       1021 TCTGGTCCCTTTAACCGTCCAGAAACTAG 1049
```

```
RESULT 2
US-09-896-186C-1
; Sequence 1, Application US/09896186C
; GENERAL INFORMATION:
;  APPLICANT: Joshua Z. Levin
;  APPLICANT:  Ken Phillips
;  APPLICANT:  Greg Budziszewski
;  APPLICANT:  Fred Meins
;  APPLICANT:  Zhenya Glazov
;  TITLE OF INVENTION: Methods of Controlling Gene Expression
;  FILE REFERENCE: 31481USNP
;  CURRENT APPLICATION NUMBER: US/09/896,186C
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: 60/222,202
;  PRIOR FILING DATE: 2000-08-01
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 942
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-896-186C-1

  Query Match             79.2%;  Score 831;  DB 5;  Length 942;
  Best Local Similarity   96.1%;  Pred. No. 2.2e-219;
  Matches  852;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy         42 ATGTCATCGTCAAATTGGATCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGAC 101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCATCGTCAAATTGGATCGACGACGCTTTTACAGAGGAAGAGCTTCTCGCTATCGAC 60

Qy        102 GCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCG 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCATCGAAGCTTCCTACAATTTCTCCCGTTCTTCTTCTTCTTCTTCCTCTGCTGCTCCG 120

Qy        162 ACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAAT 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCGTACAAGCTACAACCTCCGTCCATGGCCACGAGGAGGATCCAAATCAAATCCCCAAT 180

Qy        222 AATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCT 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AATATCCGTCGCCAATTGCCTCGTTCCATCACTTCTTCTACATCTTATAAACGATTTCCT 240

Qy        282 CTCTCCCGTTGCCGAGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTAT 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTCTCCCGTTGCCGAGCTAGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTAT 300

Qy        342 AGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACC 401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGCAAGACTGCTACTGAGGTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACC 360

Qy        402 AAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTT 461
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTT 420

Qy        462 AGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTAT 521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTAT 480

Qy        522 TGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAA 581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAA 540

Qy        582 GATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCAT 641
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCAT 600

Qy        642 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 701
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 660

Qy        702 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 720

Qy        762 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 780

Qy        822 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGAC 881
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTAACAACAACG 840

Qy        882 CTTCCTGATGCTGTCAGTGGCTCATAACGTGAAGGAGGAAGCTTAAA 928
              ||  | |||    |    | | ||||| |  || || | |
Db        841 AAAAACCATCTTCTCACACTCAACGACCTTGAAGCAAAAATCTCACA 887


RESULT 3
US-10-424-599-84537
; Sequence 84537, Application US/10424599
```

```
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 84537
;   LENGTH: 1253
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_47352C.1
US-10-424-599-84537

  Query Match             26.3%;  Score 276.2;  DB 8;  Length 1253;
  Best Local Similarity   66.6%;  Pred. No. 8.9e-66;
  Matches  395;  Conservative    0;  Mismatches  198;  Indels    0;  Gaps    0;

Qy        300 AGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAG 359
              | ||  || || | ||||| |||| |||||   ||||  ||||||| ||||  |  ||
Db        276 ATGACATTGCCTGTAATGAAGTTTAGTGGTCAAATTTCTTATAGCAGGACTTTTGATGCT 335

Qy        360 GTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGA 419
              || || ||    ||||   |||| |  || |||| |    || |      ||
Db        336 GTAGAGAAAGCTGCAACAAAGCTCTTACAAATTCTCCAAGAAAAAACGACCGACATGATG 395

Qy        420 ATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCG 479
                | ||    |||| || || |||||||||| ||| |  || ||||||||||||  ||
Db        396 CAAACTGCAATTGGATTTGACATTGAGTGGAAACCCACCTTCAGAAAAGGTGTTCCTCCC 455

Qy        480 GGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATT 539
              || ||||| ||      | ||||||||||  || | ||   ||||| ||||| | ||| |
Db        456 GGAAAGGTAGCAGTGATGCAGATATGTGGTGACACTAGACATTGTCATGTTCTACATCTA 515

Qy        540 TTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAG 599
              ||||||||| |||||||||| | | || | | |  |||||||| | ||| |  | |||
Db        516 ATTCATTCTGGAATCCCTCAAAATTTACAGCTTTTGCTTGAAGATCCCACAGTCTTGAAG 575

Qy        600 GTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATC 659
              || ||   ||| |||||||||||  |||||||| |||    || |||     |   | |
Db        576 GTTGGAGCTGGGATTGATGGTGATGCTGTGAAGGTTTTTAGAGATTATAACATATCTGTT 635

Qy        660 AAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATTGGTGGAGATAAAAAATGG 719
              |||| |||   |||||||||   |   || || |||||  ||||||||||| | || |||
Db        636 AAAGGTGTGACGGATCTTTCTTTTCATGCTAATCAAAAGCTTGGTGGAGATCATAAGTGG 695

Qy        720 GGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATC 779
              || ||||| ||  | ||||| | |||| | |  ||| ||||    ||||| |||| |||
Db        696 GGTCTTGCATCTTTGACTGAAAAACTTCTATCAAAACAGCTTAAAAAGCCCAACAAAATA 755

Qy        780 AGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACG 839
              || || || || ||||||  |   |  | |||||||| |||| || | || || |||||
Db        756 AGACTGGGAAATTGGGAGGCTCCTGTTTTGTCAAAGGAGCAACTAGAGTATGCTGCAACA 815

Qy        840 GATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGACCTTCCTGATGC 892
              ||||||| |||||| |||  ||||||  |||   |||| || || || || ||
Db        816 GATGCTTTTGCTTCTTGGTGTCTTTATCAGGCGATTAAAGATCTCCCGGACGC 868


RESULT 4
US-10-389-048-9441
; Sequence 9441, Application US/10389048
; GENERAL INFORMATION:
;  APPLICANT: Havukkala, Ilkka
;  APPLICANT:  Shenk, Michael A.
;  TITLE OF INVENTION: Polynucleotides, Materials Incorporating
;  TITLE OF INVENTION:  Them, and Methods for Using Them
;  FILE REFERENCE: 11000.1041U1c1
;  CURRENT APPLICATION NUMBER: US/10/389,048
;  CURRENT FILING DATE:  2003-03-13
;  NUMBER OF SEQ ID NOS: 25129
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9441
;   LENGTH: 672
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-10-389-048-9441

  Query Match             17.3%;  Score 181.2;  DB 8;  Length 672;
  Best Local Similarity   60.9%;  Pred. No. 1.4e-39;
  Matches  313;  Conservative    0;  Mismatches  198;  Indels    3;  Gaps    1;

Qy        382 TTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATA 441
              || | ||      |  || ||| || |  || |   |    ||   |  ||||||| || |
Db          1 TTTTGAAGAAAATCAATGCCATGAAAAGTGTAATGGATCAAGTCATAATTGGCTTTGACA 60

Qy        442 TTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGA 501
              ||||||  || || |  |||  |||||||||||  ||  |||||   ||      | ||||
Db         61 TTGAGTATAGGCCCACATTTCAAAAAGGTGTTCAACCTCGGAAGACGGCATTAATGCAGA 120

Qy        502 TATGTGTAGATAGTAATTATTGTGATGTTATGCATATTTTTCATTCTGGTATCCCTCAAA 561
              | ||||  || | | |  | ||  |||| ||||||  || ||||||||||  | ||||  |
Db        121 TTTGTGGTGACAATGACCACTGCCATGTGATGCATGTTATTCATTCTGGCGTACCTCCCA 180

Qy        562 GTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTG 621
              || |  | | | || | || ||| ||||  |  |||||||  ||    || || | |||||
Db        181 GTTTGAAGCTTATTTTAGAGGATACAACTGTATTAAAGGTTGGAGCAGGCATAGCTGGTG 240

Qy        622 ACTCTGTGAAGCTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAG 681
              |  ||  |||| |  |  | || ||    ||   | |||| | | || ||||||||||||
Db        241 ATGCTAGGAAGGTGCTTAAGGATTACAATGTATCTGTCAATGCTTTTCAGGATCTTTCAG 300

Qy        682 ATTTAGCCAACCAAAAAATTG---GTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTG 738
              |  | || || |||||  |||           | | |  ||||| ||||| || ||||| |
Db        301 AACTTGCTAATCAAAAGCTTGCCATGCCTCGTTATAGGTGGGGTCTTGCATCTCTAACCG 360

Qy        739 AGACACTTGTTTGCAAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGT 798
              ||   ||| |    ||  ||||    ||    |  |  || |||   || || ||||||
Db        361 AGGTGCTTCTAAATAAGCAGCTGAACAAATTCATGAAGATTAGGTGCGGAAATTGGGAGG 420

Qy        799 TTTATCCTCTGTCAAAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGC 858
               |  |  | | |||||  | ||   ||  || || || || || | | |||||| ||||
Db        421 CTCTTTTTTTATCAAAAGAACAACAACTTTATGCTGCCACTGACGCATTTGCTTCTTGGC 480

Qy        859 ATCTTTACAAGGTTCTTAAGGACCTTCCTGATGC 892
              |||| ||  ||||  | |||  ||| || |||||
Db        481 ATCTATATCAGGTCTTAAAGAGCCTGCCAGATGC 514


RESULT 5
US-09-531-113-12947
; Sequence 12947, Application US/09531113
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  Heck, Gregory R.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15761)B
;  CURRENT APPLICATION NUMBER: US/09/531,113
;  CURRENT FILING DATE:  2000-03-22
;  NUMBER OF SEQ ID NOS: 48629
; SEQ ID NO 12947
;   LENGTH: 579
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: jC-gmfl02220143d02a1
US-09-531-113-12947

  Query Match             13.6%;  Score 142.4;  DB 5;  Length 579;
```

```
  Best Local Similarity   64.9%;  Pred. No. 7.5e-29;
  Matches  226;  Conservative    0;  Mismatches  121;  Indels    1;  Gaps    1;

Qy        300 AGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAG 359
              | ||  || || | ||||| |||| |||||   ||||  ||||||| ||||  |  ||
Db        228 ATGACATTGCCTGTAATGAAGTTTAGTGGTCAAATTTCTTATAGCAGGACTTTTGATGCT 287

Qy        360 GTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGA 419
              || || ||    ||||   ||||  |  || ||||  |    || |     ||
Db        288 GTAGAGAAAGCTGCAACAAAGCTCTTACAAATTCTCCAAGAAAAAACGACCGACATGATG 347

Qy        420 ATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCG 479
                | ||    |||| || || |||||||||| ||| |  || | ||||||||||  ||
Db        348 CAAACTGCAATTGGATTTGACATTGAGTGGAAACCCACCTTCA-AAAAGGTGTTCCTCCC 406

Qy        480 GGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATT 539
              || ||||| ||     | ||||||||||  || | ||   ||||| ||||| | ||| |
Db        407 GGAAAGGTAGCAGTGATGCAGATATGTGGTGACACTAGACATTGTCATGTTCTACATCTA 466

Qy        540 TTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAG 599
               |||||||||| |||||||||| | | || | | |  |||||||| | ||| |  | |||
Db        467 ATTCATTCTGGAATCCCTCAAAATTTACAGCTTTTGCTTGAAGATCCCACAGTCTTGAAG 526

Qy        600 GTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTAT 647
              || ||   ||| |||||||||||  |||||| | ||||  |  | |||
Db        527 GTTGGAGCTGGGATTGATGGTGATGCTGTGAGGGTTTTAGAGAATTAT 574


RESULT 6
US-09-531-113-12947
; Sequence 12947, Application US/09531113
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  Heck, Gregory R.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15761)B
;  CURRENT APPLICATION NUMBER: US/09/531,113
;  CURRENT FILING DATE:  2000-03-22
;  NUMBER OF SEQ ID NOS: 48629
; SEQ ID NO 12947
;   LENGTH: 579
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: jC-gmfl02220143d02a1
US-09-531-113-12947

  Query Match             13.6%;  Score 142.4;  DB 6;  Length 579;
  Best Local Similarity   64.9%;  Pred. No. 7.5e-29;
  Matches  226;  Conservative    0;  Mismatches  121;  Indels    1;  Gaps    1;

Qy        300 AGGAATTTTCCAGCAATGAGGTTTGGTGGTAGGATTTTGTATAGCAAGACTGCTACTGAG 359
              | ||  || || | ||||| |||| |||||   ||||  |||||| ||||  |  ||
Db        228 ATGACATTGCCTGTAATGAAGTTTAGTGGTCAAATTTCTTATAGCAGGACTTTTGATGCT 287

Qy        360 GTTGATAAGCGAGCAATGCAGCTTATTAAAGTTCTTGATACCAAGAGAGATGAATCTGGA 419
              || || ||    ||||   ||||  |  || ||||  |    || |     ||
Db        288 GTAGAGAAAGCTGCAACAAAGCTCTTACAAATTCTCCAAGAAAAAACGACCGACATGATG 347

Qy        420 ATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCAAGTTTTAGAAAAGGTGTTCTCCCG 479
                | ||    |||| || || |||||||||| ||| |  || | ||||||||||  ||
Db        348 CAAACTGCAATTGGATTTGACATTGAGTGGAAACCCACCTTCA-AAAAGGTGTTCCTCCC 406

Qy        480 GGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGTAATTATTGTGATGTTATGCATATT 539
              || ||||| ||     | ||||||||||  || | ||   ||||| ||||| | ||| |
Db        407 GGAAAGGTAGCAGTGATGCAGATATGTGGTGACACTAGACATTGTCATGTTCTACATCTA 466

Qy        540 TTTCATTCTGGTATCCCTCAAAGTCTCCAACATCTTATTGAAGATTCAACACTTGTAAAG 599
               |||||||||| |||||||||| | | || | | |  |||||||| | ||| |  | |||
Db        467 ATTCATTCTGGAATCCCTCAAAATTTACAGCTTTTGCTTGAAGATCCCACAGTCTTGAAG 526

Qy        600 GTAGGTATTGGAATTGATGGTGACTCTGTGAAGCTTTTCCATGACTAT 647
              || ||   ||| |||||||||||  |||||| | ||||  |  | |||
Db        527 GTTGGAGCTGGGATTGATGGTGATGCTGTGAGGGTTTTAGAGAATTAT 574


RESULT 7
US-09-531-113-705
; Sequence 705, Application US/09531113
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  Heck, Gregory R.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15761)B
;  CURRENT APPLICATION NUMBER: US/09/531,113
;  CURRENT FILING DATE:  2000-03-22
;  NUMBER OF SEQ ID NOS: 48629
; SEQ ID NO 705
;   LENGTH: 278
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700909421H1
US-09-531-113-705

  Query Match             11.9%;  Score 124.6;  DB 5;  Length 278;
  Best Local Similarity   68.5%;  Pred. No. 5e-24;
  Matches  172;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        642 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 701
              || |||    |   | | |||| |||   |||||||||   |   || || |||||  ||
Db          4 GATTATAACATATCTGTTAAAGGTGTGACGGATCTTTCTTTTCATGCTAATCAAAAGCTT 63

Qy        702 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 761
              |||||||||  | || ||||| |||||  ||  | |||||  | |||| | |  ||| ||||
Db         64 GGTGGAGATCATAAGTGGGGTCTTGCATCTTTGACTGAAAAACTTCTATCAAAACAGCTT 123

Qy        762 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 821
                |||||  |||| ||| || || || || ||||||   |   |  | ||||||||  ||||
Db        124 AAAAAGCCCAACAAAATAAGACTGGGAAATTGGGAGGCTCCTGTTTTGTCAAAGGAGCAA 183

Qy        822 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGAC 881
                || | || || ||||| ||||||| ||||| |||  ||||||  |||   |||| ||
Db        184 CTAGAGTATGCTGCAACAGATGCTTTTGCTTCTTGGTGTCTTTATCAGGCGATTAAAGAT 243

Qy        882 CTTCCTGATGC 892
              || || || ||
Db        244 CTCCCGGACGC 254


RESULT 8
US-09-531-113-705
; Sequence 705, Application US/09531113
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  Heck, Gregory R.
;  APPLICANT:  La Rosa, Thomas J.
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15761)B
;  CURRENT APPLICATION NUMBER: US/09/531,113
;  CURRENT FILING DATE:  2000-03-22
;  NUMBER OF SEQ ID NOS: 48629
; SEQ ID NO 705
;   LENGTH: 278
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: 700909421H1
US-09-531-113-705

  Query Match             11.9%;  Score 124.6;  DB 6;  Length 278;
  Best Local Similarity   68.5%;  Pred. No. 5e-24;
  Matches  172;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;
```

```
Qy        642 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 701
              || |||    |   | | |||| |||   |||||||||   |   || || |||||  ||
Db          4 GATTATAACATATCTGTTAAAGGTGTGACGGATCTTTCTTTTCATGCTAATCAAAAGCTT 63

Qy        702 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 761
              ||||||||| | || ||||| ||||| ||  | ||||| | |||| | |  ||| ||||
Db         64 GGTGGAGATCATAAGTGGGGTCTTGCATCTTTGACTGAAAAACTTCTATCAAAACAGCTT 123

Qy        762 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 821
                ||||| |||| ||| || || || || ||||||  |  |  | ||||||| ||||
Db        124 AAAAAGCCCAACAAAATAAGACTGGGAAATTGGGAGGCTCCTGTTTTGTCAAAGGAGCAA 183

Qy        822 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGAC 881
              || | || || ||||| ||||||| |||||| |||  ||||||  |||   |||| ||
Db        184 CTAGAGTATGCTGCAACAGATGCTTTTGCTTCTTGGTGTCTTTATCAGGCGATTAAAGAT 243

Qy        882 CTTCCTGATGC 892
              || || || ||
Db        244 CTCCCGGACGC 254
```

RESULT 9
US-09-615-606A-75994
; Sequence 75994, Application US/09615606A
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; APPLICANT: Buehler, Robert E.
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Coombs, Brian E.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Nelson, Donald E.
; APPLICANT: Shukla, Hridayabhiranjan
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15444)C
; CURRENT APPLICATION NUMBER: US/09/615,606A
; CURRENT FILING DATE: 2000-07-13
; NUMBER OF SEQ ID NOS: 91663
; SEQ ID NO 75994
; LENGTH: 278
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700909421H1
US-09-615-606A-75994

Query Match 11.9%; Score 124.6; DB 6; Length 278;
Best Local Similarity 68.5%; Pred. No. 5e-24;
Matches 172; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```
Qy        642 GACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAACCAAAAAATT 701
              || |||    |   | | |||| |||   |||||||||   |   || || |||||  ||
Db          4 GATTATAACATATCTGTTAAAGGTGTGACGGATCTTTCTTTTCATGCTAATCAAAAGCTT 63

Qy        702 GGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGCAAAGAGCTC 761
              ||||||||| | || ||||| ||||| ||  | ||||| | |||| | |  ||| ||||
Db         64 GGTGGAGATCATAAGTGGGGTCTTGCATCTTTGACTGAAAAACTTCTATCAAAACAGCTT 123

Qy        762 CTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCAAAGCAGCAG 821
                ||||| |||| ||| || || || || ||||||  |  |  | ||||||| ||||
Db        124 AAAAAGCCCAACAAAATAAGACTGGGAAATTGGGAGGCTCCTGTTTTGTCAAAGGAGCAA 183

Qy        822 TTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTTCTTAAGGAC 881
              || | || || ||||| ||||||| |||||| |||  ||||||  |||   |||| ||
Db        184 CTAGAGTATGCTGCAACAGATGCTTTTGCTTCTTGGTGTCTTTATCAGGCGATTAAAGAT 243

Qy        882 CTTCCTGATGC 892
              || || || ||
Db        244 CTCCCGGACGC 254
```

RESULT 10
US-09-724-676-40395
; Sequence 40395, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40395
; LENGTH: 3255
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676-40395

Query Match 10.6%; Score 111; DB 6; Length 3255;
Best Local Similarity 54.1%; Pred. No. 6.5e-20;
Matches 272; Conservative 0; Mismatches 225; Indels 6; Gaps 2;

```
Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || ||||||  ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||    || || |||||      | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  |||||  | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |      |      | |  ||| |||||  |   ||   | |    ||  | |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||    ||    ||| |   | | || |||      | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932
```

RESULT 11
US-09-724-676A-40395
; Sequence 40395, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40395
; LENGTH: 3255
; TYPE: DNA

```
;   ORGANISM: Homo sapiens
US-09-724-676A-40395

  Query Match            10.6%;  Score 111;  DB 6;  Length 3255;
  Best Local Similarity   54.1%;  Pred. No. 6.5e-20;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || |||||||  ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||    || || |||||     | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  ||||| | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |    |    | |  ||| |||||  |   ||  | |    ||  |  |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||    ||    ||| |   | | || |||      | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932


RESULT 12
US-09-896-186C-17
; Sequence 17, Application US/09896186C
; GENERAL INFORMATION:
;  APPLICANT: Joshua Z. Levin
;  APPLICANT:  Ken Phillips
;  APPLICANT:  Greg Budziszewski
;  APPLICANT:  Fred Meins
;  APPLICANT:  Zhenya Glazov
;  TITLE OF INVENTION: Methods of Controlling Gene Expression
;  FILE REFERENCE: 31481USNP
;  CURRENT APPLICATION NUMBER: US/09/896,186C
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: 60/222,202
;  PRIOR FILING DATE: 2000-08-01
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
;   LENGTH: 4299
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-896-186C-17

  Query Match            10.6%;  Score 111;  DB 5;  Length 4299;
  Best Local Similarity   54.1%;  Pred. No. 7.2e-20;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || |||||||  ||||
Db        202 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 261

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||    || || |||||     | ||| | |||||   |
Db        262 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 318

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        319 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 378

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
Db        379 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 438

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  ||||| | |||||
Db        439 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 498

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |    |    | |  ||| |||||  |   ||  | |    ||  |  |
Db        499 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 558

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||    ||    ||| |   | | || |||      | ||||||  |
Db        559 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 618

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        619 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 678

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        679 TTAGAGATTTTGGATGATACTGT 701


RESULT 13
US-09-949-001-3
; Sequence 3, Application US/09949001
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL000789
;  CURRENT APPLICATION NUMBER: US/09/949,001
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/231,323
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 5208
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-001-3

  Query Match            10.6%;  Score 111;  DB 6;  Length 5208;
  Best Local Similarity   54.1%;  Pred. No. 7.6e-20;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || |||||||  ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||    || || |||||     | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
```

```
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  |||||  | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |     |     | |  ||| |||||  |   ||   | |    ||  |  |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||    ||    ||| |   | | || |||       | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932


RESULT 14
US-10-374-077-70
; Sequence 70, Application US/10374077
;   GENERAL INFORMATION:
;        APPLICANT: Fu, Ying-Hui
;                   Yu, Chang-En
;                   Oshima, Junko
;                   Mulligan, John T.
;                   Schellenberg, Gerald D.
;        TITLE OF INVENTION: ANTIBODIES AGAINST GENE PRODUCTS RELATED TO
;                            WERNER'S SYNDROME
;        NUMBER OF SEQUENCES: 209
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Seed Intellectual Property Law Group
;             STREET: 701 Fifth Avenue, Suite 6300
;             CITY: Seattle
;             STATE: Washington
;             COUNTRY: USA
;             ZIP: 98104-7092
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/374,077
;             FILING DATE: 25-Feb-2003
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Rosenman, Stephen
;             REGISTRATION NUMBER: 43,058
;             REFERENCE/DOCKET NUMBER: 100107.401D1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (206) 622-4900
;             TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 70:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 5208 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 70:
US-10-374-077-70

  Query Match             10.6%;  Score 111;  DB 8;  Length 5208;
  Best Local Similarity   54.1%;  Pred. No. 7.6e-20;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || ||||||  ||||
Db        433 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 492

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||   || || |||||     | ||| | |||||   |
Db        493 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 549

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        550 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 609

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
Db        610 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 669

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692
              ||| | | ||||| ||   | |   | ||  || ||| |||  |  |||||  | |||||
Db        670 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 729

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752
              ||||  |     |     | |  ||| |||||  |   ||   | |    ||  |  |
Db        730 AAAAAGCTGAAATGTACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 789

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812
              ||| ||||||||||    ||    ||| |   | | || |||       | ||||||  |
Db        790 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 849

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872
              || | |||  ||  || ||||| || ||||||||||||   |    | |||||     |
Db        850 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 909

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895
              |  ||    |   |||| ||||
Db        910 TTAGAGATTTTGGATGATACTGT 932


RESULT 15
US-09-949-001-9
; Sequence 9, Application US/09949001
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL000789
;  CURRENT APPLICATION NUMBER: US/09/949,001
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/231,323
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 5365
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-001-9

  Query Match             10.6%;  Score 111;  DB 6;  Length 5365;
  Best Local Similarity   54.1%;  Pred. No. 7.7e-20;
  Matches  272;  Conservative    0;  Mismatches  225;  Indels    6;  Gaps    2;

Qy        396 GATACCAAGAGAGATGAATCTGGAATAGCTTTTGTTGGCTTGGATATTGAGTGGAGACCA 455
              ||||  |  |    |  ||| |     | | | || || || || || ||||||  ||||
Db        434 GATATTAGCATGAGTCTATCAGATGGGGATGTGGTGGGATTTGACATGGAGTGGCCACCA 493

Qy        456 AGTTTTAGAAAAGGTGTTCTCCCGGGGAAGGTTGCGACTGTCCAGATATGTGTAGATAGT 515
                |  |  | |||    ||   || || |||||     | ||| | |||||   |
Db        494 TTATACAATAGAGGGAAACT---TGGCAAAGTTGCACTAATTCAGTTGTGTGTTTCTGAG 550

Qy        516 AATTATTGTGATGTTATGCATATTTTT---CATTCTGGTATCCCTCAAAGTCTCCAACAT 572
              |   | ||| |  |  | ||  ||| |      || | | | || ||  |  |  ||
Db        551 AGCAAATGTTACTTGTTCCACGTTTCTTCCATGTCAGTTTTTCCCCAGGGATTAAAAATG 610

Qy        573 CTTATTGAAGATTCAACACTTGTAAAGGTAGGTATTGGAATTGATGGTGACTCTGTGAAG 632
              |  ||||| ||  | || ||  ||||| |||| | |||||||| || ||      |||
```

```
Db        611 TTGCTTGAAAATAAAGCAGTTAAAAAGGCAGGTGTAGGAATTGAAGGAGATCAGTGGAAA 670

Qy        633 CTTTTCCATGACTATGGAGTTAGTATCAAAGATGTTGAGGATCTTTCAGATTTAGCCAAC 692

Db        671 CTTCTACGTGACTTTGATATCAAATTGAAGAATTTTGTGGAGTTGACAGATGTTGCCAAT 730

Qy        693 CAAAAAATTGGTGGAGATAAAAAATGGGGCCTTGCCTCACTAACTGAGACACTTGTTTGC 752

Db        731 AAAAAGCTGAAATGCACAGAGACCTGGAGCCTTAACAGTCTGGTTAAACACCTCTTAGGT 790

Qy        753 AAAGAGCTCCTGAAGCCAAACAGAATCAGGCTTGGGAACTGGGAGTTTTATCCTCTGTCA 812

Db        791 AAACAGCTCCTGAAAGACAAGTCTATCCGCTGTAGCAATTGGAGTAAATTTCCTCTCACT 850

Qy        813 AAGCAGCAGTTACAATACGCAGCAACGGATGCTTATGCTTCATGGCATCTTTACAAGGTT 872

Db        851 GAGGACCAGAAACTGTATGCAGCCACTGATGCTTATGCTGGTTTTATTATTTACCGAAAT 910

Qy        873 CTTAAGGACCTTCCTGATGCTGT 895

Db        911 TTAGAGATTTTGGATGATACTGT 933

Search completed: June 14, 2003, 09:27:35
Job time : 920 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 6, 2003, 10:17:10 ; Search time 36 Seconds
(without alignments)
1066.005 Million cell updates/sec

Title: US-09-896-186B-24
Perfect score: 1491
Sequence: 1 MSSSNWIDDAFTEEELLAID.....YASWHLYKVLKDLPDAVSGS 288

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1491	100.0	288	21	AAG52858	Arabidopsis thalia
2	1491	100.0	288	23	AAM50937	Arabidopsis RNase
3	1491	100.0	298	21	AAG52857	Arabidopsis thalia
4	1439.5	96.5	313	23	AAM50927	Arabidopsis RNase
5	1433	96.1	288	21	AAG24977	Arabidopsis thalia
6	1028	68.9	197	21	AAG52859	Arabidopsis thalia
7	996	66.8	197	21	AAG24978	Arabidopsis thalia
8	896	60.1	177	21	AAG24979	Arabidopsis thalia
9	352	23.6	1432	18	AAY14517	Human WRN gene pro
10	352	23.6	1432	23	AAM50935	Human Werner syndr
11	339	22.7	1401	19	AAW59454	Mouse WRN helicase
12	339	22.7	1401	20	AAW97841	Murine mWRN polype
13	330	22.1	1401	18	AAY14519	Mouse WRN gene pro
14	229.5	15.4	346	22	ABB71575	Drosophila melanog
15	212.5	14.3	583	22	ABB62653	Drosophila melanog
16	179.5	12.0	582	23	ABB91374	Herbicidally activ
17	179.5	12.0	582	23	AAM50938	Arabidopsis RNase
18	178.5	12.0	244	22	AAB80638	Environmental stre
19	151	10.1	127	22	ABB11264	Human secreted pro
20	151	10.1	147	20	AAY29336	Human secreted pro
21	151	10.1	147	22	AAU39087	Human secreted pro
22	151	10.1	147	23	ABB55796	Human polypeptide
23	151	10.1	496	22	AAB93116	Human protein sequ
24	136.5	9.2	910	23	AAM50934	Nematode 3'-5' exo
25	127.5	8.6	239	23	AAM50931	Arabidopsis RNase
26	121	8.1	505	23	AAM50939	Arabidopsis RNase
27	112.5	7.5	217	23	AAM50932	Arabidopsis RNase
28	105	7.0	625	22	ABB58154	Drosophila melanog
29	103	6.9	220	21	AAG04684	Arabidopsis thalia
30	96	6.4	706	16	AAR75647	Thermophilic bacte
31	95	6.4	161	21	AAG28363	Arabidopsis thalia
32	95	6.4	167	21	AAG28362	Arabidopsis thalia
33	94.5	6.3	416	21	AAB18185	Plasmodium falcipa
34	93	6.2	325	22	ABG17672	Novel human diagno
35	93	6.2	2224	17	AAW04254	Human Factor V. H
36	93	6.2	2224	20	AAY49564	Human lipoprotein
37	92	6.2	237	23	AAM50929	Arabidopsis RNase
38	91	6.1	451	21	AAG18364	Arabidopsis thalia
39	91	6.1	477	21	AAG18363	Arabidopsis thalia
40	89.5	6.0	346	15	AAR60653	pstS variant. Esc
41	89.5	6.0	930	22	AAU35540	Haemophilus influe
42	89.5	6.0	1078	22	ABB61252	Drosophila melanog
43	89.5	6.0	1487	22	ABB62792	Drosophila melanog
44	89	6.0	930	23	ABP40469	Staphylococcus epi
45	89	6.0	991	21	AAY83171	Cell wall protein

ALIGNMENTS

RESULT 1
AAG52858
ID AAG52858 standard; Protein; 288 AA.
XX
AC AAG52858;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 67237.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

```
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match           100.0%;  Score 1491;  DB 21;  Length 288;
  Best Local Similarity  100.0%;  Pred. No. 5.6e-151;
  Matches  288;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288


RESULT 2
AAM50937
ID   AAM50937 standard; Protein; 288 AA.
XX
AC   AAM50937;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Arabidopsis RNase D related domain (3'-5' exonuclease domain).
XX
KW   RNase D; enzyme; 3'-5' exonuclease; gene expression;
KW   gene silencing; transgenic plant.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210362-A2.
XX
PD   07-FEB-2002.
XX
PF   30-JUL-2001; 2001WO-EP08825.
XX
PR   01-AUG-2000; 2000US-222202P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA   (NOVS ) NOVARTIS FORSCHUNGSSTIFTUNG ZWEIGNIEDERL.
XX
PI   Levin JZ,  Phillips KL,  Budziszewski GJ,  Meins F,  Glazov EA;
XX
DR   WPI; 2002-217119/27.
DR   N-PSDB; ABA91805.
XX
PT   Regulating gene expression in plants for controlling gene silencing,
PT   comprises altering the transcription or translation of an endonuclease
PT   nucleotide sequence encoding a polypeptide comprising an exonuclease
PT   domain  -
XX
PS   Claim 1; Page 94-95; 102pp; English.
XX
CC   The present is that of an Arabidopsis thaliana 3'-5' exonuclease
CC   domain, specifically an RNase D related domain.  The 3'-5'
CC   exonuclease domain consists of 3 sequence motifs termed ExoI, ExoII
CC   and ExoIII.  It was identified on the basis of homology to the
CC   Arabidopsis 3'-5' exonuclease sequence given in AAM50927.  The
CC   nucleotide (see ABA91795-807) and polypeptide (see AAM50927-39)
CC   sequences for 3'-5' exonuclease domains of the invention were
CC   identified using different screening methods and search algorithms.
CC   The invention encompasses the suppression or increase of gene
CC   silencing in plants.  This is achieved by altering the expression
CC   in the plant cell of a nucleotide sequence encoding a polypeptide,
CC   such as the present sequence, having an exonuclease domain,
CC   preferably a 3'-5' exonuclease domain, and especially an RNase D
CC   related domain.  Expression of the nucleotide sequence is altered
CC   by altering its transcription or translation.  It is reduced e.g.
CC   by sense suppression, antisense suppression, homologous
CC   recombination, the use of ribozymes, dominant-negative mutants,
CC   aptamers, zinc finger proteins, double-stranded RNA, or insertional,
CC   point or deletion mutagenesis, and increased by overexpression.
CC   The methods are useful for altering or stabilizing the expression of
CC   a nucleotide sequence of interest in a plant cell or plant.
CC   Suppressing or decreasing expression of the nucleic acid molecule
CC   results in decreased levels of post transcriptional gene silencing
CC   and improved expression of genes of interest.  The 3'-5'
CC   exonuclease domain polypeptides are also useful for identifying
CC   compounds capable of altering 3'-5' exonuclease activity.
XX
SQ   Sequence   288 AA;

  Query Match           100.0%;  Score 1491;  DB 23;  Length 288;
  Best Local Similarity  100.0%;  Pred. No. 5.6e-151;
  Matches  288;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288


RESULT 3
AAG52857
ID   AAG52857 standard; Protein; 298 AA.
XX
AC   AAG52857;
```

XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 67236.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

```
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match          100.0%;  Score 1491;  DB 21;  Length 298;
  Best Local Similarity  100.0%;  Pred. No. 5.9e-151;
  Matches  288;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 70

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 130

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 190

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 250

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 298


RESULT 4
AAM50927
ID   AAM50927 standard; Protein; 313 AA.
XX
AC   AAM50927;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Arabidopsis RNase D related domain (3'-5' exonuclease domain).
XX
KW   RNase D; enzyme; 3'-5' exonuclease; gene expression; plant;
KW   gene silencing; transgenic plant.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   Domain          129..287
FT                   /label= Exonuclease
FT   Peptide         129..138
FT                   /label= ExoI
FT   Peptide         186..201
FT                   /label= ExoII
FT   Peptide         260..270
FT                   /label= ExoIII
XX
PN   WO200210362-A2.
XX
PD   07-FEB-2002.
XX
PF   30-JUL-2001; 2001WO-EP08825.
XX
PR   01-AUG-2000; 2000US-222202P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA   (NOVS ) NOVARTIS FORSCHUNGSSTIFTUNG ZWEIGNIEDERL.
XX
PI   Levin JZ,  Phillips KL,  Budziszewski GJ,  Meins F,  Glazov EA;
XX
DR   WPI; 2002-217119/27.
DR   N-PSDB; ABA91795.
XX
PT   Regulating gene expression in plants for controlling gene silencing,
PT   comprises altering the transcription or translation of an endonuclease
PT   nucleotide sequence encoding a polypeptide comprising an exonuclease
PT   domain  -
XX
PS   Claim 1; Page 70-71; 102pp; English.
XX
CC   The present is that of an Arabidopsis thaliana 3'-5' exonuclease
CC   domain (GenPept accession CAB36851), specifically an RNase D related
CC   domain.  The 3'-5' exonuclease domain consists of 3 sequence motifs
CC   termed ExoI, ExoII and ExoIII.  Nucleotide (see ABA91795-807) and
CC   polypeptide (see AAM50927-39) sequences for 3'-5' exonuclease
CC   domains were identified using different screening methods and search
CC   algorithms.  The invention encompasses the suppression or increase
CC   of gene silencing in plants.  This is achieved by altering the
CC   expression in the plant cell of a nucleotide sequence encoding a
CC   polypeptide, such as the present sequence, having an exonuclease
CC   domain, preferably a 3'-5' exonuclease domain, and especially
CC   an RNase D related domain.  Expression of the nucleotide sequence
CC   is altered by altering its transcription or translation.  It is
CC   reduced e.g. by sense suppression, antisense suppression, homologous
CC   recombination, the use of ribozymes, dominant-negative mutants,
CC   aptamers, zinc finger proteins, double-stranded RNA, or insertional,
CC   point or deletion mutagenesis, and increased by overexpression.
CC   The methods are useful for altering or stabilizing the expression of
CC   a nucleotide sequence of interest in a plant cell or plant.
CC   Suppressing or decreasing expression of the nucleic acid molecule
CC   results in decreased levels of post transcriptional gene silencing
```

```
CC   and improved expression of genes of interest.  The 3'-5'
CC   exonuclease domain polypeptides are also useful for identifying
CC   compounds capable of altering 3'-5' exonuclease activity.
XX
SQ   Sequence   313 AA;

  Query Match             96.5%;  Score 1439.5;  DB 23;  Length 313;
  Best Local Similarity   95.3%;  Pred. No. 2.1e-145;
  Matches  281;  Conservative    1;  Mismatches    4;  Indels    9;  Gaps    1;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKV---------LKDLPDAVS 286
              ||||||||||||||||||||||||||||||||||||         | ||   :|
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVTTTKNHLLTLNDLEAKIS 295


RESULT 5
AAG24977
ID   AAG24977 standard; Protein; 288 AA.
XX
AC   AAG24977;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 28856.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
```

PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 96.1%; Score 1433; DB 21; Length 288;
Best Local Similarity 96.5%; Pred. No. 9.2e-145;
Matches 278; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSLSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              |||||||||||  || ||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITCPTSSKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              |||||| |||| ||||||||||||||||||||||:||||||||||||||||||||||||||
Db        121 KRDESGRAFVGFDIEWRPSFRKGVLPGKVATVQICVDNNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              |||||||||||||||||||||||||||||||||||||||||| |:||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIDGEKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
              |||||||||||||||||||||||||||||||||||||||||||||||| |
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSDS 288
```

RESULT 6
AAG52859
ID AAG52859 standard; Protein; 197 AA.
XX
AC AAG52859;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 67238.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

```
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             68.9%;  Score 1028;  DB 21;  Length 197;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-101;
  Matches  197;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          92 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVAT 151
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVAT 60

Qy         152 VQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 211
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 VQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 120

Qy         212 LSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYAS 271
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 LSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYAS 180

Qy         272 WHLYKVLKDLPDAVSGS 288
               |||||||||||||||||
Db         181 WHLYKVLKDLPDAVSGS 197


RESULT 7
AAG24978
ID   AAG24978 standard; Protein; 197 AA.
XX
AC   AAG24978;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 28857.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
```

PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 66.8%; Score 996; DB 21; Length 197;
Best Local Similarity 97.0%; Pred. No. 3.5e-98;
Matches 191; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
Qy         92 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVAT 151
              ||||||||||||||||||||||||||||||||| |||| ||||||||||||||||||||
Db          1 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGRAFVGFDIEWRPSFRKGVLPGKVAT 60

Qy        152 VQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 211
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VQICVDNNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 120

Qy        212 LSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYAS 271
              ||||||||| |:|||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LSDLANQKIDGEKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYAS 180

Qy        272 WHLYKVLKDLPDAVSGS 288
              ||||||||||||||| |
Db        181 WHLYKVLKDLPDAVSDS 197
```

RESULT 8
AAG24979
ID AAG24979 standard; Protein; 177 AA.
XX
AC AAG24979;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 28858.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

```
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             60.1%;  Score 896;  DB 21;  Length 177;
  Best Local Similarity   96.6%;  Pred. No. 1.6e-87;
  Matches  171;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        112 MQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGI 171
              ||||||||||||||| |||| ||||||||||||||||||||||:|||||||||||||
Db          1 MQLIKVLDTKRDESGRAFVGFDIEWRPSFRKGVLPGKVATVQICVDNNYCDVMHIFHSGI 60

Qy        172 PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASL 231
              |||||||||||||||||||||||||||||||||||||||||||||||||| |:||||||||
Db         61 PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIDGEKKWGLASL 120

Qy        232 TETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        121 TETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSDS 177


RESULT 9
AAY14517
ID   AAY14517 standard; Protein; 1432 AA.
XX
AC   AAY14517;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Human WRN gene product.
XX
KW   Human; WRN; Werner's syndrome; detection; diagnosis; autosomal;
KW   recessive disorder; phenotype.
XX
OS   Homo sapiens.
XX
PN   WO9724435-A1.
XX
PD   10-JUL-1997.
XX
PF   30-DEC-1996;   96WO-US20785.
XX
PR   12-APR-1996;   96US-0632175.
PR   29-DEC-1995;   95US-0009409.
PR   29-DEC-1995;   95US-0580539.
PR   30-JAN-1996;   96US-0010835.
PR   30-JAN-1996;   96US-0594242.
XX
PA   (DARW-) DARWIN MOLECULAR CORP.
PA   (OSHI/) OSHIMA J.
XX
PI   Fu Y,  Mulligan J,  Oshima J,  Schellenberg GD,  Yu C;
XX
DR   WPI; 1997-363671/33.
DR   N-PSDB; AAX83001.
XX
PT   Isolated nucleic acid molecule encoding the WRN gene product -
PT   useful for detection and treatment of Werner's syndrome, and related
PT   diseases
XX
PS   Claim 10; Fig 2B; 153pp; English.
XX
CC   This sequence represents the human WRN gene product which is associated
CC   with Werner's syndrome.  The products can be used for the detection and
CC   treatment of Werner's syndrome (WS), an autosomal recessive disorder
CC   with a complex phenotype, as well as related diseases.
XX
SQ   Sequence   1432 AA;

  Query Match             23.6%;  Score 352;  DB 18;  Length 1432;
  Best Local Similarity   37.7%;  Pred. No. 7.6e-28;
  Matches   75;  Conservative   41;  Mismatches   75;  Indels    8;  Gaps    3;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              : | : | | |:||  |::    :       |     |    || |:|| | : :| | |
Db         43 DLPFLEFTGSIVYSYDASDCSFLSE------DISMSLSDGDVVGFDMEWPPLYNRGKL-G 95

Qy        148 KVATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSI 206
              ||| :|:||  : | : |:    : || |: |:|:  : | |:||:||  ||  |: : :
Db         96 KVALIQLCVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKL 155

Qy        207 KDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
              |:  :|:|:||:|:    : | | || : |: |:|||   ||  ||  :||:: |  ||||
Db        156 KNFVELTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLYAAT 215

Qy        267 DAYASWHLYKVLKDLPDAV 285
              |||| : :|: |: | | |
Db        216 DAYAGFIIYRNLEILDDTV 234


RESULT 10
AAM50935
ID   AAM50935 standard; Protein; 1432 AA.
XX
AC   AAM50935;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Human Werner syndrome protein (3'-5' exonuclease domain).
XX
KW   Enzyme; 3'-5' exonuclease; human; gene expression;
KW   Werner syndrome protein; gene silencing; transgenic plant.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          78..219
FT                   /label= Exonuclease
FT   Peptide         78..87
FT                   /label= ExoI
FT   Peptide         135..150
FT                   /label= ExoII
FT   Peptide         209..219
FT                   /label= ExoIII
XX
PN   WO200210362-A2.
```

```
XX
PD   07-FEB-2002.
XX
PF   30-JUL-2001; 2001WO-EP08825.
XX
PR   01-AUG-2000; 2000US-222202P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA   (NOVS ) NOVARTIS FORSCHUNGSSTIFTUNG ZWEIGNIEDERL.
XX
PI   Levin JZ,  Phillips KL,  Budziszewski GJ,  Meins F,  Glazov EA;
XX
DR   WPI; 2002-217119/27.
DR   N-PSDB; ABA91803.
XX
PT   Regulating gene expression in plants for controlling gene silencing,
PT   comprises altering the transcription or translation of an endonuclease
PT   nucleotide sequence encoding a polypeptide comprising an exonuclease
PT   domain  -
XX
PS   Claim 14; Page 86-91; 102pp; English.
XX
CC   The present is that of human Werner syndrome protein (GenPept
CC   accession AAF06162), which includes an N-terminal 3'-5' exonuclease
CC   domain.  The protein was identified in a homology screening with an
CC   Arabidopsis thaliana 3'-5' exonuclease domain (see AAM50927).  The
CC   3'-5' exonuclease domains consists of 3 sequence motifs, termed
CC   ExoI, ExoII and ExoIII.  Nucleotide (see ABA91795-807) and
CC   polypeptide (see AAM50927-39) sequences for 3'-5' exonuclease
CC   domains were identified using different screening methods and search
CC   algorithms.  The invention encompasses the suppression or increase
CC   of gene silencing in plants.  This is achieved by altering the
CC   expression in the plant cell of a nucleotide sequence encoding a
CC   polypeptide, such as the present sequence, having an exonuclease
CC   domain, preferably a 3'-5' exonuclease domain, and especially
CC   an RNase D related domain.  Expression of the nucleotide sequence
CC   is altered by altering its transcription or translation.  It is
CC   reduced e.g. by sense suppression, antisense suppression, homologous
CC   recombination, the use of ribozymes, dominant-negative mutants,
CC   aptamers, zinc finger proteins, double-stranded RNA, or insertional,
CC   point or deletion mutagenesis, and increased by overexpression.
CC   The methods are useful for altering or stabilizing the expression of
CC   a nucleotide sequence of interest in a plant cell or plant.
CC   Suppressing or decreasing expression of the nucleic acid molecule
CC   results in decreased levels of post transcriptional gene silencing
CC   and improved expression of genes of interest.  The 3'-5'
CC   exonuclease domain polypeptides are also useful for identifying
CC   compounds capable of altering 3'-5' exonuclease activity.
XX
SQ   Sequence   1432 AA;

  Query Match             23.6%;  Score 352;  DB 23;  Length 1432;
  Best Local Similarity   37.7%;  Pred. No. 7.6e-28;
  Matches   75;  Conservative   41;  Mismatches   75;  Indels    8;  Gaps    3;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              : | : | | |:||  |::    :       |     |    || |:|| | : :| | |
Db         43 DLPFLEFTGSIVYSYDASDCSFLSE------DISMSLSDGDVVGFDMEWPPLYNRGKL-G 95

Qy        148 KVATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSI 206
              ||| :|:||  : | : |:    : || |: |:|:  : | |:||:||  ||  |: : :
Db         96 KVALIQLCVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKL 155

Qy        207 KDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
              |:  :|:|:||:|:   : | | || : |: |:|||   ||  ||  :||:: |  ||||
Db        156 KNFVELTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLYAAT 215

Qy        267 DAYASWHLYKVLKDLPDAV 285
              |||| : :|: |: | | |
Db        216 DAYAGFIIYRNLEILDDTV 234


RESULT 11
AAW59454
ID   AAW59454 standard; Protein; 1401 AA.
XX
AC   AAW59454;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Mouse WRN helicase protein.
XX
KW   Werner's syndrome; WRN; helicase; murine; diagnosis.
XX
OS   Mus sp.
XX
PN   JP10146188-A.
XX
PD   02-JUN-1998.
XX
PF   15-NOV-1996;   96JP-0304721.
XX
PR   15-NOV-1996;   96JP-0304721.
XX
PA   (EIJI-) EIJIN KENKYUSHO KK.
XX
DR   WPI; 1998-369863/32.
DR   N-PSDB; AAV35114, AAV35115.
XX
PT   New mouse gene encoding protein - used for diagnosis of human
PT   Werner's syndrome
XX
PS   Claim 1; Page 18-22; 30pp; Japanese.
XX
CC   This sequence represents a novel mouse WRN helicase protein which is
CC   associated with Werner's syndrome. The mouse WRN gene is useful for
CC   the diagnosis of human Werner's syndrome.
XX
SQ   Sequence   1401 AA;

  Query Match             22.7%;  Score 339;  DB 19;  Length 1401;
  Best Local Similarity   36.6%;  Pred. No. 1.8e-26;
  Matches   74;  Conservative   42;  Mismatches   72;  Indels   14;  Gaps    5;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              | | : | | |:||  |::       | : :  :  :  :  || |:|| | ::    ||
Db         37 NLPFLEFPGSIVYSYEASD----CSFLSEDISMRLSDGDV--VGFDMEWPPIYK----PG 86

Qy        148 K---VATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG 203
              |   || :|:||  | | : ||    : || |: |:|: :: | |:||:||  ||  |:
Db         87 KRSRVAVIQLCVSENKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRDFD 146

Qy        204 VSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQY 263
              | ::   :|:|:||:|:   : | |  | : :: |:|||   ||  ||  :||:: |  |
Db        147 VKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLY 206

Qy        264 AATDAYASWHLYKVLKDLPDAV 285
              |||||||   :|: | :| | |
Db        207 AATDAYAGLIIYQKLGNLGDTV 228


RESULT 12
AAW97841
ID   AAW97841 standard; Protein; 1401 AA.
XX
AC   AAW97841;
XX
DT   07-JUN-1999  (first entry)
XX
DE   Murine mWRN polypeptide (WRN homologue).
XX
KW   mWRN gene; WRN gene; homologue; Werner's syndrome; mouse;
KW   life span; ageing.
XX
OS   Mus sp.
XX
```

PN WO9906543-A1.
XX
PD 11-FEB-1999.
XX
PF 03-AUG-1998; 98WO-US16081.
XX
PR 04-AUG-1997; 97US-0054629.
XX
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Guarente LP, Lombard DB, Sinclair DA;
XX
DR WPI; 1999-153773/13.
DR N-PSDB; AAX24302.
XX
PT Identifying agents that inhibit replication and accumulation of DNA
PT circles - as potential agents for increasing the life span of
PT animals
XX
PS Claim 13; Page 69-72; 78pp; English.
XX
CC This is the amino acid sequence of a polypeptide encoded by mWRN
CC see AAX24302), the murine homologue of the human WRN gene associated
CC with Werner's syndrome. A claimed assay for identifying a compound
CC which extends life span comprises administering a compound to be
CC tested to a mammal with a suppressed level of mWRN, and identifying
CC slowing of at least one of the normal phenotypes of ageing in the
CC mammal. A compound identified by this method inhibits the
CC replication and/or accumulation of rDNA circles in a cell. The
CC assay can be performed using a mWRN knockout mouse (also claimed).
CC mWRN nucleic acids can be used to produce mWRN polypeptides and as
CC diagnostic probes and primers. mWRN polypeptides can be used to
CC raise antibodies (for inhibiting activity of mWRN or for determining
CC it in tissues). The assay is based on the observation that in yeast
CC cells, accumulation of ribosomal circular DNA is responsible for
CC age-related enlargement and fragmentation of the nucleolus. A
CC mutation in the yeast WRN homologue SGS1 causes premature ageing,
CC suggesting a common ageing mechanism in all eukaryotes. Insight
CC into the ageing process in model systems can provide insight into
CC ageing in humans.
XX
SQ Sequence 1401 AA;

Query Match 22.7%; Score 339; DB 20; Length 1401;
Best Local Similarity 36.6%; Pred. No. 1.8e-26;
Matches 74; Conservative 42; Mismatches 72; Indels 14; Gaps 5;

```
Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              | | : | | |:|| |:: | : : : : || |:|| | :: ||
Db         37 NLPFLEFPGSIVYSYEASD----CSFLSEDISMRLSDGDV--VGFDMEWPPIYK----PG 86

Qy        148 K---VATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG 203
              |   || :|:|| | | : ||    : || |: |:|: :: | |:||:|| || |:
Db         87 KRSRVAVIQLCVSENKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRDFD 146

Qy        204 VSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQY 263
              | ::   :|:|:||:|:   : | |  | : :: |:||| || || :||:: |  |
Db        147 VKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLY 206

Qy        264 AATDAYASWHLYKVLKDLPDAV 285
              |||||||   :|: | :| | |
Db        207 AATDAYAGLIIYQKLGNLGDTV 228
```

RESULT 13
AAY14519
ID AAY14519 standard; Protein; 1401 AA.
XX
AC AAY14519;
XX
DT 31-AUG-1999 (first entry)
XX
DE Mouse WRN gene product.
XX
KW Human; WRN; Werner's syndrome; detection; diagnosis; autosomal;
KW recessive disorder; phenotype.
XX
OS Mus musculus.
XX
PN WO9724435-A1.
XX
PD 10-JUL-1997.
XX
PF 30-DEC-1996; 96WO-US20785.
XX
PR 12-APR-1996; 96US-0632175.
PR 29-DEC-1995; 95US-0009409.
PR 29-DEC-1995; 95US-0580539.
PR 30-JAN-1996; 96US-0010835.
PR 30-JAN-1996; 96US-0594242.
XX
PA (DARW-) DARWIN MOLECULAR CORP.
PA (OSHI/) OSHIMA J.
XX
PI Fu Y, Mulligan J, Oshima J, Schellenberg GD, Yu C;
XX
DR WPI; 1997-363671/33.
DR N-PSDB; AAX83004.
XX
PT Isolated nucleic acid molecule encoding the WRN gene product -
PT useful for detection and treatment of Werner's syndrome, and related
PT diseases
XX
PS Claim 10; Fig 6; 153pp; English.
XX
CC This sequence represents the mouse WRN gene product. The corresponding
CC human gene product (AAY14517) is associated with Werner's syndrome. The
CC products can be used for the detection and treatment of Werner's
CC syndrome (WS), an autosomal recessive disorder with a complex phenotype,
CC as well as related diseases.
XX
SQ Sequence 1401 AA;

Query Match 22.1%; Score 330; DB 18; Length 1401;
Best Local Similarity 36.0%; Pred. No. 1.7e-25;
Matches 72; Conservative 43; Mismatches 71; Indels 14; Gaps 5;

```
Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              | | : | | |:|| |:: | : : : : || |:|| | :: ||
Db         37 NLPFLEFPGSIVYSYEASD----CSFLSEDISMRLSDGDV--VGFDMEWPPIYK----PG 86

Qy        148 K---VATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG 203
              |   || :|:|| | | : ||    : || |: |:|: :: | |:||:|| || |:
Db         87 KRSRVAVIQLCVSESKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRDFD 146

Qy        204 VSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQY 263
              | ::   :|:|:||:|:   : | |  | : :: |:||| || || :||:: |  |
Db        147 VKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLY 206

Qy        264 AATDAYASWHLYKVLKDLPD 283
              |||||||   :|: | :| |
Db        207 AATDAYAGLIIYQKLGNLGD 226
```

RESULT 14
ABB71575
ID ABB71575 standard; Protein; 346 AA.
XX
AC ABB71575;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 41517.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

```
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15678.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 41517; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   346 AA;

  Query Match             15.4%;  Score 229.5;  DB 22;  Length 346;
  Best Local Similarity   26.9%;  Pred. No. 1.2e-15;
  Matches   66;  Conservative   43;  Mismatches  123;  Indels   13;  Gaps    5;

Qy         41 TVQATTSVHGHEEDPNQIPNNIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILY 100
              |   |  :   ||:|   |     :| ||  |          : :    | ::: | | |
Db         68 TPDVTEKLAMEEENP---PKRRSSRLTRSTRSMAEDGSPSPEKEKPEKLPFIKYKGAIKY 124

Qy        101 SKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNY 160
              : ::   |  ::: :: ::||  :  :  |:||  ||: |  ||| | :|||||
Db        125 FTESQDIAASADDVLQWVEKQKDE--VVPMAFDMEWPFSFQTG--PGKSAVIQICVDEKC 180

Qy        161 CDVMHIFH-SGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDY-----GVSIKDVEDLSD 214
              | :  : :   :| :|  ||     :   |: |  |  ||  |:        |:   ||
Db        181 CYIYQLTNVKKLPAALVALINHPKVRLHGVNIKNDFRKLARDFPEVTAEPLIEKCVDLGL 240

Qy        215 LANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHL 274
               |:      :| |  ||  :  | : |  ::|:  |   || : || ||| | |    :
Db        241 WCNEVCETGGRWSLERLTNFIAKKAMDKSKKVRMSKWHVIPLDENQLMYAAIDVYIGQVI 300

Qy        275 YKVLK 279
              |: |:
Db        301 YRELE 305


RESULT 15
ABB62653
ID   ABB62653 standard; Protein; 583 AA.
XX
AC   ABB62653;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 14751.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06756.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 14751; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   583 AA;

  Query Match             14.3%;  Score 212.5;  DB 22;  Length 583;
  Best Local Similarity   33.5%;  Pred. No. 1.7e-13;
  Matches   61;  Conservative   23;  Mismatches   65;  Indels   33;  Gaps    6;

Qy        109 KRAMQLIKVLDTKRDESGIAFVGLDIEW-------RPSFRKGVLPGKVATVQICVDSNYC 161
              |   |  |||             | | ||       ||          || :|:      |
Db         71 KNHCQTFKVL------------GFDCEWITVGGSRRP----------VALLQLSSHRGLC 108

Qy        162 DVMHIFH-SGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKI 220
               :  : |   ||| |: |:|| :::|||:    |::|| |||||  :    ||  |    :
Db        109 ALFRLCHMKQIPQDLRELLEDDSVIKVGVAPQEDAMKLSHDYGVGVASTLDLRFLC--VM 166

Qy        221 GGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLY-KVLK 279
               | |  ||  |::| :   | |  |:   |||   |  :|| ||| ||  :  :| |: :
Db        167 AGHKPEGLGKLSKTHLNYTLDKHWRLACSNWEAKTLEPKQLDYAANDALMAVAIYQKLCR 226

Qy        280 DL 281
              ||
Db        227 DL 228


Search completed: June  6, 2003, 10:23:34
Job time : 38 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 6, 2003, 10:21:45 ; Search time 19 Seconds
(without alignments)
1457.195 Million cell updates/sec

Title: US-09-896-186B-24
Perfect score: 1491
Sequence: 1 MSSSNWIDDAFTEEELLAID.....YASWHLYKVLKDLPDAVSGS 288

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1439.5	96.5	313	2	T05256	hypothetical prote
2	339.5	22.8	1436	2	T14895	DNA helicase 1 - A
3	339	22.7	1401	2	T30247	Werner syndrome pr
4	330	22.1	1401	2	T17452	Werner syndrome pr
5	179.5	12.0	582	2	H96604	probable 3'-5' exo
6	158	10.6	123	2	E84752	hypothetical prote
7	136.5	9.2	910	2	S40930	hypothetical prote
8	127.5	8.6	239	2	A84777	hypothetical prote
9	112.5	7.5	217	2	T02548	hypothetical prote
10	101	6.8	876	2	T19246	hypothetical prote
11	97.5	6.5	445	2	E72674	hypothetical prote
12	96.5	6.5	488	2	F97039	hypothetical prote
13	95	6.4	508	1	A36121	aromatase (EC 1.14
14	95	6.4	4589	2	T14914	dynein beta heavy
15	94.5	6.3	416	2	C71620	protein with Egl-1
16	93.5	6.3	1525	1	GNWVS5	genome polyprotein
17	93	6.2	2224	1	KFHU5	coagulation factor
18	93	6.2	3079	1	RGBYI2	probable GTPase-ac
19	92	6.2	237	2	H84732	hypothetical prote
20	92	6.2	758	2	T39210	3-isopropylmalate
21	91.5	6.1	473	2	S36553	L2 protein - human
22	91.5	6.1	781	2	T41551	hypothetical prote
23	91	6.1	503	2	S13912	aromatase (EC 1.14
24	90	6.0	472	2	T47436	protein kinase-lik
25	89.5	6.0	930	2	E64098	DNA-directed DNA p
26	89.5	6.0	1324	2	T01508	mismatch repair en
27	88	5.9	269	2	I50504	nk2.2 protein - ze
28	88	5.9	317	2	A97511	hypothetical prote
29	88	5.9	317	2	AH2729	agmatinase [import
30	88	5.9	417	2	E82095	gamma-glutamyl pho
31	88	5.9	709	2	S40926	hypothetical prote
32	88	5.9	784	2	C88558	protein ZK1098.3 [
33	87.5	5.9	334	2	T03556	probable D-ribose-
34	87.5	5.9	422	2	E96753	hypothetical prote
35	87.5	5.9	547	2	T40342	signal recognition
36	87.5	5.9	589	2	E88492	protein T07E3.1 [i
37	87	5.8	178	2	E72450	hypothetical prote
38	87	5.8	392	2	T44362	acetyl-CoA C-acety
39	87	5.8	579	2	I40371	methyltransferase
40	87	5.8	3137	1	A37797	collagen alpha 3(V
41	86.5	5.8	346	1	BYECPR	phosphate-repressi
42	86.5	5.8	360	2	T00882	hypothetical prote
43	86.5	5.8	470	2	T10193	glycerol-3-phospha
44	86.5	5.8	496	2	S61966	hypothetical prote
45	86.5	5.8	593	2	B89844	hypothetical prote

ALIGNMENTS

RESULT 1
T05256
hypothetical protein F18A5.260 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C;Accession: T05256
R;Bevan, M.; Weber, N.; Grueninger, D.; Schmidheini, T.; Bancroft, I.; Mewes, H.W.; M
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15405
A;Accession: T05256
A;Molecule type: DNA
A;Residues: 1-313 <BEV>
A;Cross-references: EMBL:AL035528
A;Experimental source: cultivar Columbia; BAC clone F18A5
C;Genetics:
A;Map position: 4
A;Introns: 86/1; 143/1; 186/3; 239/3
A;Note: F18A5.260

Query Match 96.5%; Score 1439.5; DB 2; Length 313;
Best Local Similarity 95.3%; Pred. No. 5.6e-117;
Matches 281; Conservative 1; Mismatches 4; Indels 9; Gaps 1;

```
Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKV---------LKDLPDAVS 286
              |||||||||||||||||||||||||||||||||||||         | ||   :|
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVTTTKNHLLTLNDLEAKIS 295
```

RESULT 2
T14895
DNA helicase 1 - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14895
R;Yan, H.; Chen, C.Y.; Kobayashi, R.; Newport, J.
Nature Genet. 19, 375-378, 1998

A;Title: Replication focus-forming activity 1 and the Werner syndrome gene product.
A;Reference number: Z18255; MUID:98361165; PMID:9697700
A;Accession: T14895
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1436 <YAN>
A;Cross-references: EMBL:AF067418; NID:g3420290; PID:g3420291; PIDN:AAC63512.1
C;Genetics:
A;Gene: FFA-1

Query Match 22.8%; Score 339.5; DB 2; Length 1436;
Best Local Similarity 35.0%; Pred. No. 8.5e-21;
Matches 82; Conservative 43; Mismatches 84; Indels 25; Gaps 7;

```
Qy         61 NIRRQLPRSITSSTSYKRF---PLSRCRAR----NFPAMRFGGRILYSKTATEV-----D 108
              :::|:|| ::      |      | |:      | | |:| | |:|| : :      |
Db          3 SLQRKLPEWMSVKQQEDRIDDAKKSFCKKNILEDNLPFMKFNGSIVYSYESNDCSLLSED 62

Qy        109 KRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIF- 167
              |:  |         |  :  :| |||| | : ||   |||| :|:|| |  : ||
Db         63 IRSSLL---------EEDV--LGFDIEWPPVYTKG-KTGKVALIQVCVSEKKCYLFHISP 110

Qy        168 HSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWG 227
              :| |: |: |:|| :: |||:||:||   ||  || : :|   :||::||||:   :||
Db        111 MAGFPKGLKRLLEDESVRKVGVGIEGDQWKLMSDYELKLKGFIELSEMANQKLRCKEKWT 170

Qy        228 LASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDL 281
                | : |  ::| | |  ||: : |:: |  |||||||   :|| |: :
Db        171 FNGLIKHLFKEQLYKRKSYRCSNWDIFLLTEDQKLYAATDAYAGLLIYKKLEGM 224
```

RESULT 3
T30247
Werner syndrome protein typeI - mouse
N;Alternate names: Wrn typeI protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T30247
R;Imamura, O.; Ichikawa, K.; Yamabe, Y.; Goto, M.; Sugawara, M.; Furuichi, Y.
Genomics 41, 298-300, 1997
A;Title: Cloning of a mouse homologue of the human Werner syndrome gene and assignment t
A;Reference number: Z20785; MUID:97288537; PMID:9143515
A;Accession: T30247
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1401 <IMA>
A;Cross-references: EMBL:D86526; NID:g2130972; PIDN:BAA20269.1; PID:g2130973
A;Experimental source: strain BALB/c; testis/spleen
C;Genetics:
A;Gene: WRN typeI
A;Map position: 8A4

Query Match 22.7%; Score 339; DB 2; Length 1401;
Best Local Similarity 36.6%; Pred. No. 9.1e-21;
Matches 74; Conservative 42; Mismatches 72; Indels 14; Gaps 5;

```
Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              | | : | | |:||  |::       | : :  :  :  :  || |:|| | ::    ||
Db         37 NLPFLEFPGSIVYSYEASD----CSFLSEDISMRLSDGDV--VGFDMEWPPIYK----PG 86

Qy        148 K---VATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG 203
              |   || :|:||  | | : ||    : || |: |:|: :: | |:||:||  ||  |:
Db         87 KRSRVAVIQLCVSENKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRDFD 146

Qy        204 VSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQY 263
              | ::   :|:|:||:|:   : | |  | : :: |:|||   ||  ||  :||:: |  |
Db        147 VKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLY 206

Qy        264 AATDAYASWHLYKVLKDLPDAV 285
              |||||||   :|: | :| | |
Db        207 AATDAYAGLIIYQKLGNLGDTV 228
```

RESULT 4
T17452
Werner syndrome protein - mouse
N;Alternate names: Wrn protein
C;Species: Mus musculus (house mouse)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T17452
R;Paeper, B.W.; Gayle, M.; Brady, W.; Swartz, A.; Gillett, L.A.; Alisch, R.S.; Mullig
submitted to the EMBL Data Library, September 1998
A;Description: Genomic structure of the human Werner's gene and cloning of its mouse
A;Reference number: Z18794
A;Accession: T17452
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1401 <PAE>
A;Cross-references: EMBL:AF091215; NID:g3885837; PID:g3885838; PIDN:AAC78077.1
C;Genetics:
A;Gene: Wrn

Query Match 22.1%; Score 330; DB 2; Length 1401;
Best Local Similarity 36.0%; Pred. No. 5.5e-20;
Matches 72; Conservative 43; Mismatches 71; Indels 14; Gaps 5;

```
Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              | | : | | |:||  |::       | : :  :  :  :  || |:|| | ::    ||
Db         37 NLPFLEFPGSIVYSYEASD----CSFLSEDISMRLSDGDV--VGFDMEWPPIYK----PG 86

Qy        148 K---VATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG 203
              |   || :|:||  | | : ||    : || |: |:|: :: | |:||:||  ||  |:
Db         87 KRSRVAVIQLCVSESKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRDFD 146

Qy        204 VSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQY 263
              | ::   :|:|:||:|:   : | |  | : :: |:|||   ||  ||  :||:: |  |
Db        147 VKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLY 206

Qy        264 AATDAYASWHLYKVLKDLPD 283
              |||||||   :|: | :| |
Db        207 AATDAYAGLIIYQKLGNLGD 226
```

RESULT 5
H96604
probable 3'-5' exonuclease [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H96604
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H96604
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-582 <STO>
A;Cross-references: GB:AE005173; NID:g11094727; PIDN:AAG29662.1; GSPDB:GN00141
C;Genetics:
A;Gene: F14G9.8
A;Map position: 1

Query Match 12.0%; Score 179.5; DB 2; Length 582;
Best Local Similarity 28.2%; Pred. No. 1.8e-07;
Matches 53; Conservative 36; Mismatches 66; Indels 33; Gaps 7;

```
Qy        126 GIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSN--YCDVMHIFHSG---IPQSLQHLIE 180
              |   ||:| ||:|:: ||    ||: :||  |:     |:: :::     :  | |:::
Db        377 GCRVVGIDCEWKPNYIKGSKQNKVSIMQIGSDTKIFILDLIKLYNDASEILDNCLSHILQ 436
```

```
Qy        181 DSTLVKV--------------GIGIDGDSVKLFHDYGVSIKDVE------DLSDLANQKI 220
              : :|:               |     |  :|   ||  :|  |      |: :: |:
Db        437 SKSTLKLVSLTEDYPDHKLSSGYNFQCDIKQLALSYG-DLKCFERYDMLLDIQNVFNEPF 495

Qy        221 GGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKD 280
              |    ||| ||: ::   | |    |  :||  |||: ||:||| ||    |::: ::|
Db        496 G-----GLAGLTKKILGVSLNKTR--RNSDWEQRPLSQNQLEYAALDAAVLIHIFRHVRD 548

Qy        281 LPDAVSGS 288
              |   | |
Db        549 HPPHDSSS 556


RESULT 6
E84752
hypothetical protein At2g34110 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: E84752
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: E84752
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-123 <STO>
A;Cross-references: GB:AE002093; NID:g2342731; PIDN:AAB67629.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g34110
A;Map position: 2

  Query Match             10.6%;  Score 158;  DB 2;  Length 123;
  Best Local Similarity   79.5%;  Pred. No. 1.6e-06;
  Matches   31;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        190 GIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGL 228
              ||||| |||||:|||||||||:|||| |:||||| | |:
Db         82 GIDGDYVKLFHEYGVSIKDVENLSDLPNRKIGGDLKNGV 120


RESULT 7
S40930
hypothetical protein ZK1098.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1997
C;Accession: S40930
R;Thomas, K.
submitted to the EMBL Data Library, February 1992
A;Reference number: S40923
A;Accession: S40930
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-910 <THO>
A;Cross-references: EMBL:Z22176; NID:g297978; PID:g297986
C;Genetics:
A;Introns: 64/1; 336/2; 382/2; 447/2; 681/2; 810/1; 852/2

  Query Match              9.2%;  Score 136.5;  DB 2;  Length 910;
  Best Local Similarity   24.8%;  Pred. No. 0.0017;
  Matches   77;  Conservative   34;  Mismatches  110;  Indels   89;  Gaps   13;

Qy         12 TEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPRSIT 71
              ||::   |||| :     :|||| |:| |: |   |
Db        333 TEKQ---IDAICFAIYLGIASSSS-------------------YQLP-NVMRDFFRQPD 368

Qy         72 SSTSYKRFPLSRCRARNFPAMRFGGRILY-----------SKTATEVDKRAMQLIKVLD 119
              |     :  | | :    |     |  |:             || :|::     |   :
Db        369 SKLREAKELLVRRKTLQVP---LNGEQLFVFENERRTQIHMVKTESEMN----YLCSEIK 421
```

```
Qy        120 TKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQI----------CVD---SNYCDVMHI 166
              :  ||    :|| | ||:||    |   |:| :|:          ||:   :|  |
Db        422 SLSDEPAPVYVGFDSEWKPSNLTAVHDSKIAIIQLFFKNCVWLVDCVELEKANMAD---- 477

Qy        167 FHSGIPQSLQHLIEDSTLVKVGIGIDGD-----------SVKLFHDYGVSIKDVEDLSD 214
                  :    |  || :  ||  :  |            |:|:        |:  ||
Db        478 --DWWQKFASRLFGDSPVKVVGFDMRNDLDAMATIPALKSSMKI-----EDTKNAFDLKR 530

Qy        215 LANQKIGGD--------KKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
              ||      |        | : || ||  |:  || |    :  ||:  || |:|: |||
Db        531 LAENVCDIDMEILELPKKTFKLADLTHYLLGLELDKTE--QCSNWQCRPLRKKQIVYAAL 588

Qy        267 DAYASWHLYK 276
              ||      :|
Db        589 DAVVVVETFK 598


RESULT 8
A84777
hypothetical protein At2g36110 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84777
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <STO>
A;Cross-references: GB:AE002093; NID:g4678223; PIDN:AAD26968.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g36110
A;Map position: 2

  Query Match              8.6%;  Score 127.5;  DB 2;  Length 239;
  Best Local Similarity   25.1%;  Pred. No. 0.0017;
  Matches   55;  Conservative   33;  Mismatches   90;  Indels   41;  Gaps   10;

Qy         62 IRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKT-ATEVDKRAMQLIKVLDT 120
              |||:: :       |  |              || |:: : |  |   :| :  |:
Db          4 IRRRIQKRHIHENRYIDF-------------FGERLIVTVTHTTSTIRRWIHSIRFFSR 49

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFH-SGIPQSLQHLI 179
              |    :  ||||::| |    |       :|:||  | | :: : |   ||: |:  :
Db         50 LRSSHPLV-VGLDVQWTP----GGSDPPPDILQLCV-GNRCLIIQLSHCKRIPEVLRSFL 103

Qy        180 EDSTLVKVGIGIDGDSVKL---FHDYGV-SIKDVED-----LSDLANQKIGGDKKWGLAS 230
              || |:  ||:    |  ||    |   :  : |:       | : : :||
Db        104 EDETITFVGVWNSQDQGKLERFRHQLEIWRLLDIRHYLPTRLLNSSFEKI---------- 153

Qy        231 LTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAY 269
              : | |  | : |   | : ||    ||  |:  |: | |
Db        154 VEECLGYKGVRKDKEICMSNWGARSLSHDQIVQASDDVY 192


RESULT 9
T02548
hypothetical protein At2g32490 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein T26B15.5
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C;Accession: T02548; G84733
R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke
submitted to the EMBL Data Library, July 1998
A;Description: Arabidopsis thaliana chromosome II BAC T26B15 genomic sequence.
A;Reference number: Z14678
A;Accession: T02548
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
```

```
A;Residues: 1-217 <ROU>
A;Cross-references: EMBL:AC004681; NID:g3298532; PID:g3298537
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84733
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-217 <STO>
A;Cross-references: GB:AE002093; NID:g3298537; PIDN:AAC25931.1; GSPDB:GN00139
C;Genetics:
A;Gene: T26B15.5; At2g32490
A;Map position: 2
A;Introns: 25/1

  Query Match             7.5%;  Score 112.5;  DB 2;  Length 217;
  Best Local Similarity   25.4%;  Pred. No. 0.03;
  Matches   47;  Conservative   38;  Mismatches   77;  Indels   23;  Gaps   10;

Qy         94 FGGRILYSKTAT-EVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATV 152
              || |:: : | |  | :| :  |: :   |    :  ||| ::| | |     |  :  :
Db         31 FGERLIVTVTHTPSVIRRWIHSIRFVSRLRLSHPLV-VGLGVQWTP--RGSDPPPDI--L 85

Qy        153 QICVDSNYCDVMHIFH-SGIPQSLQHLIEDSTLVKVGI--GIDGDSVKLFHDYGVSIKDV 209
              |:|| :  | :: : |   :|  |:  :|| |:  ||:    | | :: || : : |  :
Db         86 QLCVGTR-CLIIQLSHCKYVPDVLRSFLEDQTITFVGVWNSQDKDKLERFH-HQLDIWRL 143

Qy        210 EDLSDLANQKIGGDKKWGLASLTETLV-----CKELLKPNRIRLGNWEFYPLSKQQLQYA 264
                :    :  :       |:|  ||:|      : : |   : : ||    ||  |:  |
Db        144 VHIRHYLHPLL-------LSSSFETIVKVYLGHEGVTKDKELCMSNWGARSLSHDQIVQA 196

Qy        265 ATDAY 269
              : | |
Db        197 SHDVY 201


RESULT 10
T19246
hypothetical protein C14A4.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19246
R;McMurray, A.
submitted to the EMBL Data Library, June 1995
A;Reference number: Z19097
A;Accession: T19246
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-876 <WIL>
A;Cross-references: EMBL:Z49909; PIDN:CAA90108.1; GSPDB:GN00020; CESP:C14A4.4
A;Experimental source: clone C14A4
C;Genetics:
A;Gene: CESP:C14A4.4
A;Map position: 2
A;Introns: 80/2; 557/2; 731/3; 833/1

  Query Match             6.8%;  Score 101;  DB 2;  Length 876;
  Best Local Similarity   21.1%;  Pred. No. 2;
  Matches   64;  Conservative   60;  Mismatches  125;  Indels   54;  Gaps   15;

Qy         15 ELLAIDAIEASYNFSRSSSSSSSAAP----------TVQATTSVHGHEEDPNQIPNNIRR 64
              ::: ::  : :||     || :| ||           ::  | :  |::| :      ||
Db        183 QMVVLEKPQKTYNI----SSDNSQAPFSSKLTVKHHAIEKRTGIVLHDDDESG-----RR 233

Qy         65 QLPRSITSS-------TSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKV 117
                : | :           | :   : ::   ::|       |  :  ::   | |
Db        234 DWISAETETEEEHPYIAEILHFKVPEAQLKSAECLKFTALKDTPLTMIDTKEKLEALTKT 293

Qy        118 LDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICV-DSNY-CDVMHIF-HSGIPQS 174
                |:: ::        :|:|      :  | |    :||   | ::  |   |: | |:
Db        294 LNSVKE------FAVDLEHHQ--MRSYL-GLTCLIQISTRDEDFIIDPFPIWDHVGM--- 341

Qy        175 LQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTET 234
              |      :  ::||  | | | : |  |||| : :: | : :| :|:    |: || |  |
Db        342 LNEPFANPRILKVFHGSDSDVLWLQRDYGVHVVNLFD-TYVAMKKL-KYPKFSLAYL--T 397

Qy        235 LVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVL---------KDLPDAV 285
              |   :::   : :| :|   ||    : ||  | :   : | :|         ||| :
Db        398 LRFADVVLDKQYQLADWRARPLRNAMINYAREDTHYLLYSYDMLREQLLKQDTKDLANVY 457

Qy        286 SGS 288
              | |
Db        458 SES 460


RESULT 11
E72674
hypothetical protein APE0818 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: E72674
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: E72674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-445 <KAW>
A;Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79797.1; PID:d1043583; PID:g
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0818

  Query Match             6.5%;  Score 97.5;  DB 2;  Length 445;
  Best Local Similarity   22.5%;  Pred. No. 1.6;
  Matches   65;  Conservative   36;  Mismatches   83;  Indels  105;  Gaps   15;

Qy         33 SSSSSAAPTV-----QATTSVHGHE-----EDPNQIPNNIRRQLPRSITSSTSYKRFPLS 82
              | :|:|||:       ::|  | |:|       :|  :|   ||||   :       |   |
Db          5 SKNSTAAPSSSHAEGESTADVLGYEGVEVLPEPKDVPQEARRQLLNLLAERGKVKPSTLG 64

Qy         83 RCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFV----------GL 132
                ||  |  || | |          :    :: :  | |  | :|| |           |
Db         65 VSRA-YFYQMRRGLR--------PISDSILERLLELATDDDLAGIPFFAPYVDYQRVKGY 115

Qy        133 DI--------EWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTL 184
              |:        ||  :          |: :: :||          :   |: |
Db        116 DVDRLVRLVAEWARA--------NPASAKVFLDS-----------LSAELERL------ 149

Qy        185 VKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKI-GGDKKWGLASLTETLVCKELLKP 243
                  |: | ::|      || : | :       ::  ||   | |
Db        150 -----GLVGKAIK------VSERHVREFEGYLEARVRSGDMDPGTAG------------- 185

Qy        244 NRIR--------LGNWEFYPLSKQQL-----QYAATDAYASWHLYKVLK 279
              :|:|         ||    | |:|| |     :| |:    : |:|| ||
Db        186 DRLRYLRMALEELG----YVLTKQALRGLIRRYQASQPGVADHIYKSLK 230


RESULT 12
F97039
hypothetical protein CAC1132 [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: F97039
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: F97039
```

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-488 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK79105.1; PID:g15024051; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1132

  Query Match             6.5%;  Score 96.5;  DB 2;  Length 488;
  Best Local Similarity   22.1%;  Pred. No. 2.2;
  Matches   34;  Conservative   33;  Mismatches   46;  Indels   41;  Gaps    5;

Qy        103 TATEVDKRAMQLI-KVLDTKRDESGIAFVGLDIEW--------RPSFRKGVLPGKVATVQ 153
              |   |  :| :|: ::|: |  |: :      :||        : |:|:  :
Db        177 TPAVVSAKATKLVAELLNIKESENEMVKKINPLEWEEKILAELKKSYRRTEI-------- 228

Qy        154 ICVDSNYCDVM-----------------HIFHSGIPQSLQHLIEDSTLVKVGIGIDGDS 195
                : ||  ||                  :|| :|:   :    :| :||::   :|||:
Db        229 --AELNYNQVMLSLRSQLKDNLGDCYWNFYIFSNGVDFLVMENYDDGSLVRIDFTVDGDN 286

Qy        196 VKLFHDYGVSIK----DVEDLSDLANQKIGGDKK 225
              | : : | |:||    : |: :::| :|    ||
Db        287 VVILNVYPVTIKYDSIENEEENEMAEKKKNPKKK 320


RESULT 13
A36121
aromatase (EC 1.14.14.-) cytochrome P450 19 - rat
N;Alternate names: cytochrome P450 arom
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C;Accession: A36121; S16901; S26817
R;Hickey, G.J.; Krasnow, J.S.; Beattie, W.G.; Richards, J.S.
Mol. Endocrinol. 4, 3-12, 1990
A;Title: Aromatase cytochrome P450 in rat ovarian granulosa cells before and after lutei
matase cDNA and 5' genomic DNA.
A;Reference number: A36121; MUID:90220647; PMID:2157976
A;Accession: A36121
A;Molecule type: mRNA
A;Residues: 1-508 <HIC>
A;Cross-references: EMBL:M33986; NID:g203804; PIDN:AAA41044.1; PID:g203805
R;Lephart, E.D.; Peterson, K.G.; Noble, J.F.; George, F.W.; McPhaul, M.J.
Mol. Cell. Endocrinol. 70, 31-40, 1990
A;Title: The structure of cDNA clones encoding the aromatase P-450 isolated from a rat L
 line.
A;Reference number: S16901; MUID:90255798; PMID:2340950
A;Accession: S16901
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2,'FQ',5-9,'Q',11-87,'C',89-100,'L',102-122,'VVH',126,'HAR',130-187,'L',18
492,'SPRNS' <LEP>
C;Genetics:
A;Gene: CYP19
C;Superfamily: human cytochrome P450 CYP19; cytochrome P450 homology
C;Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F;303-459/Domain: cytochrome P450 homology <P45>
F;437/Binding site: heme iron (Cys) (axial ligand) #status predicted

  Query Match             6.4%;  Score 95;  DB 1;  Length 508;
  Best Local Similarity   23.1%;  Pred. No. 3.1;
  Matches   42;  Conservative   25;  Mismatches   41;  Indels   74;  Gaps    9;

Qy         93 RFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEW---RPSFRKGVL-PGK 148
              |||              || :| | :      |:|| |      |   || | | :  ||
Db        115 RFG------------SKRGLQCIGM-----HENGIIFNNNPSLWRTVRPFFMKALTGPGL 157

Qy        149 VATVQICVD--------------SNYCDVMHIFHSGIPQSLQHLIED-STLVKVGIGID 192
              :  |::||:               | | ||: :        ::|:: | |  : :|| :|
Db        158 IRMVEVCVESIKQHLDRLGDVTDNSGYVDVVTL--------MRHIMLDTSNTLFLGIPLD 209

Qy        193 GDSVK-----------------------LFHDYGVSIKDVED----LSDLANQKIGGD 223
              |:                         |:  |  |:||::|    | :   ||:
Db        210 ESSIVKKIQGYFNAWQALLIKPNIFFKISWLYRKYERSVKDLKDEIEILVEKKRQKVSSA 269

Qy        224 KK 225
              :|
Db        270 EK 271


RESULT 14
T14914
dynein beta heavy chain - Tetrahymena thermophila
C;Species: Tetrahymena thermophila
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000
C;Accession: T14914
R;Lincoln, L.M.; Gibson, T.M.; Asai, D.J.; Forney, J.D.
submitted to the EMBL Data Library, June 1999
A;Description: A gene knockout reveals that dynein beta heavy chain is required in Te
A;Reference number: Z18264
A;Accession: T14914
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4589 <LIN>
A;Cross-references: EMBL:AF072878; NID:g5209335; PID:g3309593; PIDN:AAC26117.1
C;Genetics:
A;Gene: DYH4
A;Genetic code: SGC5
A;Introns: 286/3; 666/3; 4360/2; 4535/3
C;Superfamily: dynein heavy chain, ciliary
C;Keywords: P-loop

  Query Match             6.4%;  Score 95;  DB 2;  Length 4589;
  Best Local Similarity   24.9%;  Pred. No. 62;
  Matches   68;  Conservative   40;  Mismatches   97;  Indels   68;  Gaps   14;

Qy         42 VQATTSV---------HGHEEDPNQIPNNIR---RQLPRSITS-----STSYKRFPLSRC 84
              ||| |||             ::|  :  : |:   |:|   ||       | | || |
Db        463 VQAFTSVSYDIMNIAEKKFDDDFYEFRSKIKELERRLASVITQGFDDYDTIYGRFKL--- 519

Qy         85 RARNFPAMRFGGRILYSKTATEVDKRAMQLIKVL--DTKRDES----GIAFVGLDIEWRP 138
                       | | :     | |::|: : |:::   | |: :|    |  ||    | |
Db        520 ------LDNFEGLLTRPIIADELEKKHIVLLEMYKQDLKQVQSIFLEGKQFVDSMHENAP 573

Qy        139 SFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIED--STLVKVGIGID---- 192
                 |    :|     :  |                  :||:  |::   | ::| ||
Db        574 LFLN--MPPIAGALTWC----------------KSLRDRIQEPIEKLAQLGQGITEREE 614

Qy        193 -GDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCK---ELLK----PN 244
               |  ||:     |||| ||   |: :|   |       | :||:||   :|:|    |:
Db        615 YKDVQKLYTSITKSIKDYEDQKILSWEKEVEDSSQD--KLKQTLLCKDENDLIKVNFDPS 672

Qy        245 RIR-LGNWEFYPLSKQQLQYAATDAYASWHLYK 276
              :| |   ::: | : ::   | | |    |:
Db        673 LVRLLKEVKYFLLLRLEVPTTAKDIYTKAETYR 705


RESULT 15
C71620
protein with Egl-like 3'-5' exonucl. domain PFB0215c - malaria parasite (Plasmodium f
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: C71620
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: C71620
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-416 <GAR>
A;Cross-references: GB:AE001381; GB:AE001362; NID:g3845124; PIDN:AAC71832.1; PID:g384
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0215c
```

```
  Query Match             6.3%;  Score 94.5;  DB 2;  Length 416;
  Best Local Similarity   21.9%;  Pred. No. 2.6;
  Matches   40;  Conservative   35;  Mismatches   69;  Indels   39;  Gaps    7;

Qy        106 EVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNY----- 160
              |:      |:|  : |        :|||||     : |:    |: :||| :  |
Db         60 EIKNGNYNLLKDFNMK-------MIGLDIEGYKIGKYGI----VSIIQICYEDIYIFDIY 108

Qy        161 -CDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQK 219
               || :::| : |    : ::|   ::||       |   |:: | : :|:: |     |
Db        109 KCDNVYLFINYI----KDILECDDIIKVTHDCREDCSILYNQYNIHLKNILDTQVAYNLL 164

Qy        220 IGGDKKWGLASLTETLVCKELLKP-------------NRIRLGNWEFY--PLSKQQLQYA 264
              :  : :   : |  :   :|||                | | |: :   |: |: : ||
Db        165 LKNNNNY---TNTYQISYDDLLKKYLFINNNHKIYFHKMITLDNYIYLKRPIMKELISYA 221

Qy        265 ATD 267
               |
Db        222 IQD 224


Search completed: June  6, 2003, 10:24:39
Job time : 22 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 6, 2003, 10:16:45 ; Search time 10 Seconds
(without alignments)
1194.518 Million cell updates/sec

Title: US-09-896-186B-24
Perfect score: 1491
Sequence: 1 MSSSNWIDDAFTEEELLAID.....YASWHLYKVLKDLPDAVSGS 288

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	352	23.6	1432	1	WRN_HUMAN	Q14191 homo sapien
2	339	22.7	1401	1	WRN_MOUSE	O09053 mus musculu
3	136.5	9.2	910	1	YO68_CAEEL	P34607 caenorhabdi
4	95	6.4	508	1	CPV1_RAT	P22443 rattus norv
5	93.5	6.3	1525	1	POLG_STEVM	P09732 s genome po
6	93	6.2	2224	1	FA5_HUMAN	P12259 homo sapien
7	93	6.2	3079	1	IRA2_YEAST	P19158 saccharomyc
8	92	6.2	758	1	LEU2_SCHPO	O14289 schizosacch
9	91.5	6.1	473	1	VL2_HPV03	P36744 human papil
10	91	6.1	503	1	CPV1_MOUSE	P28649 mus musculu
11	89.5	6.0	930	1	DPO1_HAEIN	P43741 haemophilus
12	89.5	6.0	1324	1	MSH6_ARATH	O04716 arabidopsis
13	89.5	6.0	1487	1	BLM_DROME	Q9vgi8 drosophila
14	88	5.9	269	1	NK2B_BRARE	Q90481 brachydanio
15	88	5.9	416	1	PROA_VIBCH	Q9kpt9 vibrio chol
16	88	5.9	784	1	YO63_CAEEL	P34603 caenorhabdi
17	87	5.8	579	1	MTC1_BACST	P43423 bacillus st
18	87	5.8	3137	1	CA36_CHICK	P15989 gallus gall
19	86.5	5.8	346	1	PSTS_ECOLI	P06128 escherichia
20	86.5	5.8	470	1	PLSB_CUCSA	Q39639 cucumis sat
21	86	5.8	552	1	FMR2_APLCA	P08021 aplysia cal
22	85.5	5.7	220	1	DEOC_MYCPI	P47722 mycoplasma
23	85.5	5.7	329	1	IDI2_STRPY	Q9a095 streptococc
24	85.5	5.7	1394	1	E75B_DROME	P17672 drosophila
25	85.5	5.7	2298	1	CU05_HUMAN	Q9y3r5 homo sapien
26	84.5	5.7	480	1	UGDH_SOYBN	Q96558 glycine max
27	84.5	5.7	830	1	GYRA_CLOAB	P94605 clostridium
28	83.5	5.6	777	1	YASB_SCHPO	Q10146 schizosacch
29	83	5.6	237	1	TRPC_THEVO	Q979v9 thermoplasm
30	83	5.6	1129	1	PHYB_SOLTU	P34094 solanum tub
31	82	5.5	229	1	RPE_CHLPN	Q9z8z9 chlamydia p
32	82	5.5	492	1	6PGD_SCHPO	P78812 schizosacch
33	82	5.5	540	1	CH60_THEBR	Q60024 thermoanaer
34	81.5	5.5	462	1	LEU2_LISIN	Q92a26 listeria in
35	81.5	5.5	636	1	GYRB_THEMA	P77993 thermotoga
36	81.5	5.5	1002	1	YEMA_DROME	P25992 drosophila
37	81	5.4	309	1	CC2B_ARATH	P25859 arabidopsis
38	81	5.4	459	1	Y819_PYRHO	O58549 pyrococcus
39	81	5.4	560	1	VAOX_PENSI	P56216 penicillium
40	81	5.4	864	1	CHEA_BORBU	Q44737 borrelia bu
41	81	5.4	885	1	PMC2_HUMAN	Q01780 homo sapien
42	80.5	5.4	399	1	RRPP_PI4HA	P22044 human parai
43	80.5	5.4	451	1	NFS1_MOUSE	Q9z1j3 mus musculu
44	80.5	5.4	473	1	VL2_HPV28	P50799 human papil
45	80.5	5.4	543	1	PROA_LEGPN	P21347 legionella

ALIGNMENTS

RESULT 1
WRN_HUMAN
ID WRN_HUMAN STANDARD; PRT; 1432 AA.
AC Q14191;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Werner syndrome helicase.
GN WRN OR RECQL2 OR RECQ3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96181115; PubMed=8602509;
RA Yu C.-E., Oshima J., Fu Y.-H., Wijsman E.M., Hisama F., Alisch R.,
RA Matthews S., Nakura J., Miki T., Ouais S., Martin G.M., Mulligan J.,
RA Schellenberg G.D.;
RT "Positional cloning of the Werner's syndrome gene.";
RL Science 272:258-262(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Paeper B.W., Gayle M., Brady W., Swartz A., Gillett L.A., Alisch R.S.,
RA Mulligan J., Galas D., Fu Y.-H.;
RT "Genomic structure of the human Werner's gene and cloning of the
RT mouse homolog.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=98284027; PubMed=9618508;
RA Marciniak R.A., Lombard D.B., Johnson F.B., Guarente L.;
RT "Nucleolar localization of the Werner syndrome protein in human
RT cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:6887-6892(1998).
RN [4]
RP REPEATS.
RX MEDLINE=99160561; PubMed=10049920;
RA Kusano K., Berres M.E., Engels W.R.;
RT "Evolution of the RECQ family of helicases: a Drosophila homolog,
RT Dmblm, is similar to the human Bloom syndrome gene.";
RL Genetics 151:1027-1039(1999).
RN [5]
RP REVIEW ON VARIANTS.
RX MEDLINE=99235545; PubMed=10220139;
RA Moser M.J., Oshima J., Monnat R.J. Jr.;
RT "WRN mutations in Werner syndrome.";
RL Hum. Mutat. 13:271-279(1999).
RN [6]
RP VARIANT ARG-1367.
RX MEDLINE=97173161; PubMed=9021029;
RA Ye L., Miki T., Nakura J., Oshima J., Kamino K., Rakugi H.,
RA Ikegami H., Higaki J., Edland S.D., Martin G.M., Ogihara T.;
RT "Association of a polymorphic variant of the Werner helicase gene with
RT myocardial infarction in a Japanese population.";
RL Am. J. Med. Genet. 68:494-498(1997).

```
RN   [7]
RP   ERRATUM.
RA   Ye L., Miki T., Nakura J., Oshima J., Kamino K., Rakugi H.,
RA   Ikegami H., Higaki J., Edland S.D., Martin G.M., Ogihara T.;
RL   Am. J. Med. Genet. 70:103-103(1997).
RN   [8]
RP   VARIANTS ILE-387 AND LEU-1074.
RX   MEDLINE=98111850; PubMed=9450180;
RA   Meisslitzer C., Ruppitsch W., Weirich-Schwaiger H., Weirich H.G.,
RA   Jabkowsky J., Klein G., Schweiger M., Hirsch-Kauffmann M.;
RT   "Werner syndrome: characterization of mutations in the WRN gene in an
RT   affected family.";
RL   Eur. J. Hum. Genet. 5:364-370(1997).
RN   [9]
RP   VARIANT ILE-387.
RA   Vidal V., Bay J.-O., Champomier F., Grancho M., Beauville L.,
RA   Glowaczower C., Lemery D., Ferrara M., Bignon Y.-J.;
RT   "The 1396del A mutation and a missense mutation or a rare polymorphism
RT   of the WRN gene detected in a French Werner family with a severe
RT   phenotype and a case of an unusual vulvar cancer.";
RL   Hum. Mutat. 11:413-414(1998).
RN   [10]
RP   VARIANTS ALA-324 AND ARG-1367.
RX   MEDLINE=99167244; PubMed=10069711;
RA   Castro E., Ogburn C.E., Hunt K.E., Tilvis R., Louhija J.,
RA   Penttinen R., Erkkola R., Panduro A., Riestra R., Piussan C.,
RA   Deeb S.S., Wang L., Edland S.D., Martin G.M., Oshima J.;
RT   "Polymorphisms at the Werner locus: I. Newly identified polymorphisms,
RT   ethnic variability of 1367Cy/Arg, and its stability in a population
RT   of Finnish centenarians.";
RL   Am. J. Med. Genet. 82:399-403(1999).
CC   -!- FUNCTION: ESSENTIAL FOR THE FORMATION OF DNA REPLICATION FOCAL
CC       CENTERS; STABLY ASSOCIATES WITH FOCI ELEMENTS GENERATING BINDING
CC       SITES FOR RP-A. EXHIBITS A MAGNESIUM-DEPENDENT ATP-DEPENDENT DNA-
CC       HELICASE ACTIVITY. MAY BE INVOLVED IN THE CONTROL OF GENOMIC
CC       STABILITY (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC   -!- DISEASE: DEFECTS IN WRN ARE THE CAUSE OF WERNER SYNDROME (WS); A
CC       RARE AUTOSOMAL RECESSIVE PROGEROID SYNDROME CHARACTERIZED BY THE
CC       PREMATURE ONSET OF MULTIPLE AGE-RELATED DISORDERS, INCLUDING
CC       ATHEROSCLEROSIS, CANCER, NON-INSULIN-DEPENDENT DIABETES MELLITUS
CC       (NIDDM), OCULAR CATARACTS AND OSTEOPOROSIS. THE MAJOR CAUSE OF
CC       DEATH (AT A MEDIAN AGE OF 47) IS MYOCARDIAL INFARCTION (MI).
CC   -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECQ SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 HRDC DOMAIN.
CC   -!- DATABASE: NAME=WRN; NOTE=WRN mutation db (Warner disease);
CC       WWW="http://www.pathology.washington.edu/werner/ws_wrn.html".
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/WRNID284.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L76937; AAC41981.1; -.
DR   EMBL; AF091214; AAC63361.1; -.
DR   EMBL; AF181897; AAF06162.1; -.
DR   EMBL; AF181896; AAF06162.1; JOINED.
DR   Genew; HGNC:12791; WRN.
DR   MIM; 604611; -.
DR   MIM; 277700; -.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002121; HRDC.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004589; RecQ.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00570; HRDC; 1.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   SMART; SM00474; 35EXOc; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00341; HRDC; 1.
DR   TIGRFAMs; TIGR00614; recq; 1.
KW   Hydrolase; Helicase; ATP-binding; DNA-binding; Nuclear protein;
KW   Polymorphism; Repeat.
FT   DOMAIN      424    477       2 X 27 AA TANDEM REPEATS OF H-L-S-P-N-D-
FT                                N-E-N-D-T-S-Y-V-I-E-S-D-E-D-L-E-M-E-M-L-
FT                                K.
FT   REPEAT      424    450       1.
FT   REPEAT      451    477       2.
FT   DOMAIN      507    510       POLY-GLU.
FT   DOMAIN     1150   1229       HRDC.
FT   NP_BIND     571    578       ATP (BY SIMILARITY).
FT   SITE        668    671       DEAH BOX.
FT   VARIANT     324    324       T -> A.
FT                                /FTId=VAR_006904.
FT   VARIANT     387    387       M -> I.
FT                                /FTId=VAR_006905.
FT   VARIANT    1074   1074       F -> L.
FT                                /FTId=VAR_007903.
FT   VARIANT    1367   1367       C -> R (ASSOCIATED WITH A HIGHER RISK OF
FT                                MYOCARDIAL INFARCTION).
FT                                /FTId=VAR_006906.
SQ   SEQUENCE   1432 AA;  162494 MW;  DF02C0059F7B62EB CRC64;

  Query Match             23.6%;  Score 352;  DB 1;  Length 1432;
  Best Local Similarity   37.7%;  Pred. No. 1.2e-21;
  Matches   75;  Conservative   41;  Mismatches   75;  Indels    8;  Gaps    3;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              : | : | | |:||  |::    :       |     |    || |:|| | : :| | |
Db         43 DLPFLEFTGSIVYSYDASDCSFLSE------DISMSLSDGDVVGFDMEWPPLYNRGKL-G 95

Qy        148 KVATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSI 206
             ||| :|:||  : | : |:    : || |: |:|:  : | |:||:||  ||  |: : :
Db         96 KVALIQLCVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKL 155

Qy        207 KDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
             |:  :|:|:||:|:   : | | || : |: |:|||   ||  ||  :||:: |  ||||
Db        156 KNFVELTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLYAAT 215

Qy        267 DAYASWHLYKVLKDLPDAV 285
             |||| : :|: |: | | |
Db        216 DAYAGFIIYRNLEILDDTV 234


RESULT 2
WRN_MOUSE
ID   WRN_MOUSE      STANDARD;      PRT;  1401 AA.
AC   O09053; O09050; Q9Z242;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Werner syndrome helicase homolog.
GN   WRN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Testis, and Spleen;
RX   MEDLINE=97288537; PubMed=9143515;
RA   Imamura O., Ichikawa K., Yamabe Y., Goto M., Sugawara M.,
RA   Furuichi Y.;
RT   "Cloning of a mouse homologue of the human Werner syndrome gene and
RT   assignment to 8A4 by fluorescence in situ hybridization.";
RL   Genomics 41:298-300(1997).
RN   [2]
RP   SUBCELLULAR LOCATION.
```

```
RX   MEDLINE=98284027; PubMed=9618508;
RA   Marciniak R.A., Lombard D.B., Johnson F.B., Guarente L.;
RT   "Nucleolar localization of the Werner syndrome protein in human
RT   cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:6887-6892(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Paeper B.W., Gayle M., Brady W., Swartz A., Gillett L.A., Alisch R.S.,
RA   Mulligan J., Galas D., Fu Y.-H.;
RT   "Genomic structure of the human Werner's gene and cloning of its mouse
RT   homolog.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY BE INVOLVED IN THE CONTROL OF GENOMIC STABILITY.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECQ SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 HRDC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D86527; BAA20270.1; -.
DR   EMBL; D86526; BAA20269.1; -.
DR   EMBL; AF091215; AAC78077.1; -.
DR   MGD; MGI:109635; Wrn.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002121; HRDC.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004589; RecQ.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00570; HRDC; 1.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   SMART; SM00474; 35EXOc; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00341; HRDC; 1.
DR   TIGRFAMs; TIGR00614; recq; 1.
KW   Helicase; ATP-binding; Nuclear protein.
FT   NP_BIND     535    542       ATP (BY SIMILARITY).
FT   SITE        632    635       DEAH BOX.
FT   DOMAIN     1115   1194       HRDC.
FT   DOMAIN     1387   1390       POLY-SER.
FT   CONFLICT    101    101       N -> S (IN REF. 3).
FT   CONFLICT    228    228       V -> A (IN REF. 3).
FT   CONFLICT    250    250       L -> S (IN REF. 3).
FT   CONFLICT    452    452       M -> V (IN REF. 3).
FT   CONFLICT    459    459       K -> T (IN REF. 3).
FT   CONFLICT    468    468       C -> R (IN REF. 3).
FT   CONFLICT    619    619       K -> Q (IN REF. 3).
FT   CONFLICT    800    800       Q -> K (IN REF. 3).
FT   CONFLICT   1021   1021       L -> S (IN REF. 3).
FT   CONFLICT   1145   1145       A -> T (IN REF. 3).
FT   CONFLICT   1181   1182       VG -> LE (IN REF. 3).
FT   CONFLICT   1252   1252       V -> A (IN REF. 3).
FT   CONFLICT   1308   1308       I -> L (IN REF. 3).
FT   CONFLICT   1356   1356       V -> A (IN REF. 3).
SQ   SEQUENCE   1401 AA;  157256 MW;  949060992467FB8C CRC64;

  Query Match             22.7%;  Score 339;  DB 1;  Length 1401;
  Best Local Similarity   36.6%;  Pred. No. 1.5e-20;
  Matches   74;  Conservative   42;  Mismatches   72;  Indels   14;  Gaps    5;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              | | : | | |:||  |::       | : :  :  :  :  || |:|| | ::    ||
Db         37 NLPFLEFPGSIVYSYEASD----CSFLSEDISMRLSDGDV--VGFDMEWPPIYK----PG 86

Qy        148 K---VATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG 203
```

```
              |   || :|:||  | | : ||    : || |: |:|: :: | |:||:||  ||  |:
Db         87 KRSRVAVIQLCVSENKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRDFD 146

Qy        204 VSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQY 263
              | ::   :|:|:||:|:   : | |  | : :: |:|||   || || :||:: | |
Db        147 VKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLY 206

Qy        264 AATDAYASWHLYKVLKDLPDAV 285
              |||||||   :|: | :| | |
Db        207 AATDAYAGLIIYQKLGNLGDTV 228


RESULT 3
YO68_CAEEL
ID   YO68_CAEEL     STANDARD;      PRT;   910 AA.
AC   P34607;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 105.6 kDa protein ZK1098.8 in chromosome III.
GN   ZK1098.8.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: TO RIBONUCLEASE D.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z22176; CAA80137.1; -.
DR   PIR; S40930; S40930.
DR   WormPep; ZK1098.8; CE00370.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   SMART; SM00474; 35EXOc; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   910 AA;  105569 MW;  5512D15423517FCD CRC64;

  Query Match             9.2%;  Score 136.5;  DB 1;  Length 910;
  Best Local Similarity   24.8%;  Pred. No. 0.0011;
  Matches   77;  Conservative   34;  Mismatches  110;  Indels   89;  Gaps   13;

Qy         12 TEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPRSIT 71
              ||::   |||| :     :|||| |:| |: |   |
Db        333 TEKQ---IDAICFAIYLGIASSSS-------------------YQLP-NVMRDFFRQPD 368

Qy         72 SSTSYKRFPLSRCRARNFPAMRFGGRILY-----------SKTATEVDKRAMQLIKVLD 119
              |    :  | | :    |     |  |:             || :|::     |  :
Db        369 SKLREAKELLVRRKTLQVP---LNGEQLFVFENERRTQIHMVKTESEMN----YLCSEIK 421
```

```
Qy        120 TKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQI----------CVD---SNYCDVMHI 166
              :  ||    :|| | ||:||    |   |:| :|:          ||:   :|  |
Db        422 SLSDEPAPVYVGFDSEWKPSNLTAVHDSKIAIIQLFFKNCVWLVDCVELEKANMAD---- 477

Qy        167 FHSGIPQSLQHLIEDSTLVKVGIGIDGD-----------SVKLFHDYGVSIKDVEDLSD 214
                    :    |  || :  ||  :  |            |:|:         |:  ||
Db        478 --DWWQKFASRLFGDSPVKVVGFDMRNDLDAMATIPALKSSMKI-----EDTKNAFDLKR 530

Qy        215 LANQKIGGD--------KKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
              ||     |          | : || ||  |:  || |    :  ||:  || ||:|: |||
Db        531 LAENVCDIDMEILELPKKTFKLADLTHYLLGLELDKTE--QCSNWQCRPLRKKQIVYAAL 588

Qy        267 DAYASWHLYK 276
              ||      :|
Db        589 DAVVVVETFK 598


RESULT 4
CPV1_RAT
ID   CPV1_RAT       STANDARD;      PRT;   508 AA.
AC   P22443;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Cytochrome P450 19 (Aromatase) (EC 1.14.14.1) (CYPXIX) (Estrogen
DE   synthetase) (P-450AROM).
GN   CYP19 OR AROM.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=90220647; PubMed=2157976;
RA   Hickey G.J., Krasnow J.S., Beattie W.G., Richards J.S.;
RT   "Aromatase cytochrome P450 in rat ovarian granulosa cells before and
RT   after luteinization: adenosine 3',5'-monophosphate-dependent and
RT   independent regulation. Cloning and sequencing of rat aromatase cDNA
RT   and 5' genomic DNA.";
RL   Mol. Endocrinol. 4:3-12(1990).
CC   -!- FUNCTION: CATALYZES THE FORMATION OF AROMATIC C18 ESTROGENS FROM
CC       C19 ANDROGENS.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33986; AAA41044.1; -.
DR   PIR; A36121; A36121.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Electron transport; Oxidoreductase; Monooxygenase; Membrane;
KW   Heme.
FT   BINDING     437    437       HEME.
SQ   SEQUENCE   508 AA;  58411 MW;  C0ADFB0FD80AB352 CRC64;

  Query Match             6.4%;  Score 95;  DB 1;  Length 508;
  Best Local Similarity   23.1%;  Pred. No. 1.6;
  Matches   42;  Conservative   25;  Mismatches   41;  Indels   74;  Gaps    9;

Qy         93 RFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEW---RPSFRKGVL-PGK 148
              |||             || :| | :      |:|| |      |   || | | :  ||
Db        115 RFG-----------SKRGLQCIGM-----HENGIIFNNNPSLWRTVRPFFMKALTGPGL 157

Qy        149 VATVQICVD--------------SNYCDVMHIFHSGIPQSLQHLIED-STLVKVGIGID 192
              :  |::||:               | | ||: :        ::|:: | |  : :|| :|
Db        158 IRMVEVCVESIKQHLDRLGDVTDNSGYVDVVTL--------MRHIMLDTSNTLFLGIPLD 209

Qy        193 GDSVK-----------------------LFHDYGVSIKDVED----LSDLANQKIGGD 223
                |:                       |:  |  |:||::|    | :   ||:
Db        210 ESSIVKKIQGYFNAWQALLIKPNIFFKISWLYRKYERSVKDLKDEIEILVEKKRQKVSSA 269

Qy        224 KK 225
              :|
Db        270 EK 271


RESULT 5
POLG_STEVM
ID   POLG_STEVM     STANDARD;      PRT;  1525 AA.
AC   P09732; Q88781; Q88782; Q88783; Q88784; Q88785; Q88786; Q88787;
AC   Q88788;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix
DE   protein (Envelope protein M); Major envelope protein E; Nonstructural
DE   proteins NS1, NS2A, NS2B; Protease/helicase (EC 3.4.21.98) (NS3)]
DE   (Fragment).
OS   St. louis encephalitis virus (strain MS1-7).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Flavivirus.
OX   NCBI_TaxID=11081;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87122172; PubMed=3027980;
RA   Trent D.W., Kinney R.M., Johnson B.J.B., Vorndam A.V., Grant J.A.,
RA   Deubel V., Rice C.M., Hahn C.;
RT   "Partial nucleotide sequence of St. Louis encephalitis virus RNA:
RT   structural proteins, NS1, NS2A, and NS2B.";
RL   Virology 156:293-304(1987).
CC   -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC       HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC       NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC       precursor polyprotein, commonly with Asp or Glu in the P6
CC       position, Cys or Thr in P1 and Ser or Ala in P1'.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC       PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC       PROTEIN C AND MRNA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16614; AAA47786.1; -.
DR   PIR; A27531; GNWVS5.
DR   InterPro; IPR000069; Flavi_M.
DR   InterPro; IPR001157; Flavi_NS1.
DR   InterPro; IPR000752; Flavi_NS2A.
DR   InterPro; IPR000487; Flavi_NS2B.
DR   InterPro; IPR001122; Flavi_capsidC.
DR   InterPro; IPR000336; Flavi_glycoprotE.
DR   InterPro; IPR002535; Flavi_propep.
DR   Pfam; PF00869; Flavi_glycoprot; 1.
DR   Pfam; PF00948; Flavi_NS1; 1.
DR   Pfam; PF01002; Flavi_NS2B; 1.
DR   Pfam; PF01003; Flavi_capsid; 1.
DR   Pfam; PF01004; Flavi_M; 1.
```

```
DR   Pfam; PF01005; Flavi_NS2A; 1.
DR   Pfam; PF01570; Flavi_propep; 1.
DR   Pfam; PF02832; Flavi_glycop_C; 1.
DR   ProDom; PD001496; Flavi_NS1; 1.
DR   ProDom; PD001556; Flavi_glycoprotE; 1.
KW   Polyprotein; Glycoprotein; Core protein; Coat protein;
KW   Envelope protein; Hydrolase; Helicase; ATP-binding; Transmembrane;
KW   Nonstructural protein.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    121       CAPSID PROTEIN C.
FT   PROPEP      122    213
FT   CHAIN       214    288       ENVELOPE GLYCOPROTEIN M.
FT   CHAIN       289    789       MAJOR ENVELOPE PROTEIN E.
FT   CHAIN       790   1203       NONSTRUCTURAL PROTEIN NS1.
FT   CHAIN      1204   1368       NONSTRUCTURAL PROTEIN NS2A.
FT   CHAIN      1369   1499       NONSTRUCTURAL PROTEIN NS2B.
FT   CHAIN      1500  >1525       PROTEASE/HELICASE (NS3).
FT   TRANSMEM    108    119       POTENTIAL.
FT   TRANSMEM    253    268       POTENTIAL.
FT   TRANSMEM    274    288       POTENTIAL.
FT   TRANSMEM    751    762       POTENTIAL.
FT   TRANSMEM    768    787       POTENTIAL.
FT   TRANSMEM   1173   1188       POTENTIAL.
FT   DISULFID    291    318       BY SIMILARITY.
FT   DISULFID    348    404       BY SIMILARITY.
FT   DISULFID    362    393       BY SIMILARITY.
FT   DISULFID    380    409       BY SIMILARITY.
FT   DISULFID    478    576       BY SIMILARITY.
FT   DISULFID    593    624       BY SIMILARITY.
FT   CARBOHYD    136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    269    269       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    442    442       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    964    964       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    996    996       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1189   1189       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER    1525   1525
SQ   SEQUENCE   1525 AA;  167891 MW;  E1A373F1E511159F CRC64;

  Query Match             6.3%;  Score 93.5;  DB 1;  Length 1525;
  Best Local Similarity   22.3%;  Pred. No. 9.4;
  Matches   47;  Conservative   34;  Mismatches   85;  Indels   45;  Gaps   11;

Qy         82 SRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDE---SGIAFVGLDIE-WR 137
              :| |: :   :  || :::  |: : |         :  :| |    |  ||  |:| |:
Db        764 ARDRSISLTLLAVGGILIFLATSVQADSGC-----AISLQRRELKCGGGIFVYNDVEKWK 818

Qy        138 PSFRKGVL--PGKVATVQICVDSNYCDVM------HIFHSGIPQSLQHLIEDSTLVKVGI 189
               ::   |   |    :|    : || :       |:    | : |  : ||:  : : :
Db        819 SDYKYFPLTPTGLAHVIQEAHANGYCGIRSTSRLEHLMWENIQRELNAIFEDNE-IDLSV 877

Qy        190 GIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLG 249
              :  |  | :       :| :||  |       | ||||     |            |||
Db        878 VVQEDP-KYYKRAPRRLKKLEDELDY------GWKKWGKTLFVEP-----------RLG 918

Qy        250 NWEFY---PLSKQQLQYAATDAYASWHLYKV 277
              |  |    | :|:        |  :|: :||
Db        919 NNTFVVDGPETKE-----CPTANRAWNSFKV 944


RESULT 6
FA5_HUMAN
ID   FA5_HUMAN      STANDARD;      PRT;  2224 AA.
AC   P12259; Q14285;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Coagulation factor V precursor (Activated protein C cofactor).
GN   F5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92232668; PubMed=1567832;
RA   Cripe L.D., Moore K.D., Kane W.H.;
RT   "Structure of the gene for human coagulation factor V.";
RL   Biochemistry 31:3777-3785(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87260886; PubMed=3110773;
RA   Jenny R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
RA   Hewick R.M., Kaufman R.J., Mann K.G.;
RT   "Complete cDNA and derived amino acid sequence of human factor V.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
RN   [3]
RP   SEQUENCE OF 1-1600 FROM N.A.
RX   MEDLINE=88107560; PubMed=2827731;
RA   Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
RT   "Cloning of cDNAs coding for the heavy chain region and connecting
RT   region of human factor V, a blood coagulation factor with four types
RT   of internal repeats.";
RL   Biochemistry 26:6508-6514(1987).
RN   [4]
RP   SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
RX   MEDLINE=86313665; PubMed=3092220;
RA   Kane W.H., Davie E.W.;
RT   "Cloning of a cDNA coding for human factor V, a blood coagulation
RT   factor homologous to factor VIII and ceruloplasmin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
RN   [5]
RP   PARTIAL SEQUENCE FROM N.A.
RC   TISSUE=Fibroblast;
RX   MEDLINE=93203619; PubMed=8454869;
RA   Shen N.L.L., Fan S.-T., Pyati J., Graff R., Lapolla R.J.,
RA   Edgington T.S.;
RT   "The serine protease cofactor factor V is synthesized by
RT   lymphocytes.";
RL   J. Immunol. 150:2992-3001(1993).
RN   [6]
RP   SULFATION.
RX   MEDLINE=94264012; PubMed=8204629;
RA   Pittman D.D., Tomkinson K.N., Michnick D., Selighsohn U.,
RA   Kaufman R.J.;
RT   "Posttranslational sulfation of factor V is required for efficient
RT   thrombin cleavage and activation and for full procoagulant activity.";
RL   Biochemistry 33:6952-6959(1994).
RN   [7]
RP   SULFATION.
RX   MEDLINE=90366699; PubMed=2168225;
RA   Hortin G.L.;
RT   "Sulfation of tyrosine residues in coagulation factor V.";
RL   Blood 76:946-952(1990).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
RX   MEDLINE=20052169; PubMed=10586886;
RA   Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
RA   Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
RA   Fuentes-Prior P.;
RT   "Crystal structures of the membrane-binding C2 domain of human
RT   coagulation factor V.";
RL   Nature 402:434-439(1999).
RN   [9]
RP   VARIANT APCR GLN-534.
RX   MEDLINE=94217810; PubMed=8164741;
RA   Bertina R.M., Koeleman B.P.C., Koster T., Rosendaal F.R.,
RA   Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
RT   "Mutation in blood coagulation factor V associated with resistance to
RT   activated protein C.";
RL   Nature 369:64-67(1994).
RN   [10]
RP   VARIANT APCR GLY-334, AND VARIANT LYS-513.
RA   Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
RT   "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese.";
```

```
RL   Blood 91:1135-1139(1998).
RN   [11]
RP   VARIANT APCR THR-334.
RA   Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
RT   "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
RT   resistance to activated protein C.";
RL   Blood 91:1140-1144(1998).
CC   -!- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
CC       WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC   -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC       chain, noncovalently bound. The interaction between the two chains
CC       is calcium-dependent.
CC   -!- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
CC       REPEATS.
CC   -!- PTM: Thrombin activates factor V proteolytically to the active
CC       cofactor, factor Va (formation of a heavy chain at the N-
CC       terminus and a light chain at the C-terminus).
CC   -!- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
CC       ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
CC   -!- DISEASE: OWREN PARAHEMOPHILIA, AN HEMORRHAGIC DIASTESIS, IS DUE
CC       TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
CC       FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR). THE APCR
CC       MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGEST THAT
CC       A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
CC       IMPLANTATION.
CC   -!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC       2 PLASTOCYANIN-LIKE REPEATS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L32779; AAB59401.1; -.
DR   EMBL; L32755; AAB59401.1; JOINED.
DR   EMBL; L32756; AAB59401.1; JOINED.
DR   EMBL; L32757; AAB59401.1; JOINED.
DR   EMBL; L32758; AAB59401.1; JOINED.
DR   EMBL; L32759; AAB59401.1; JOINED.
DR   EMBL; L32760; AAB59401.1; JOINED.
DR   EMBL; L32761; AAB59401.1; JOINED.
DR   EMBL; L32762; AAB59401.1; JOINED.
DR   EMBL; L32763; AAB59401.1; JOINED.
DR   EMBL; L32764; AAB59401.1; JOINED.
DR   EMBL; L32765; AAB59401.1; JOINED.
DR   EMBL; L32766; AAB59401.1; JOINED.
DR   EMBL; L32767; AAB59401.1; JOINED.
DR   EMBL; L32768; AAB59401.1; JOINED.
DR   EMBL; L32769; AAB59401.1; JOINED.
DR   EMBL; L32770; AAB59401.1; JOINED.
DR   EMBL; L32771; AAB59401.1; JOINED.
DR   EMBL; L32772; AAB59401.1; JOINED.
DR   EMBL; L32773; AAB59401.1; JOINED.
DR   EMBL; L32774; AAB59401.1; JOINED.
DR   EMBL; L32775; AAB59401.1; JOINED.
DR   EMBL; L32776; AAB59401.1; JOINED.
DR   EMBL; L32777; AAB59401.1; JOINED.
DR   EMBL; L32778; AAB59401.1; JOINED.
DR   EMBL; M16967; AAA52424.1; -.
DR   EMBL; M14335; AAB59532.1; -.
DR   PIR; A25897; A25897.
DR   PIR; A28028; A28028.
DR   PDB; 1CZS; 26-NOV-99.
DR   PDB; 1CZT; 26-NOV-99.
DR   PDB; 1CZV; 26-NOV-99.
DR   Genew; HGNC:3542; F5.
DR   MIM; 227400; -.
DR   MIM; 134400; -.
DR   MIM; 188055; -.
DR   MIM; 227310; -.
DR   InterPro; IPR001117; Cu-oxidase.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00394; Cu-oxidase; 3.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW   Repeat; Polymorphism; Disease mutation; Thrombophilia; 3D-structure.
FT   SIGNAL       1     28
FT   CHAIN       29   2224       COAGULATION FACTOR V.
FT   CHAIN       29    737       HEAVY CHAIN.
FT   PEPTIDE    738   1573       ACTIVATION PEPTIDE (CONNECTING REGION).
FT   CHAIN      1574   2224       LIGHT CHAIN.
FT   DOMAIN       30    329       F5/8 TYPE A 1.
FT   DOMAIN       30    193       PLASTOCYANIN-LIKE 1.
FT   DOMAIN      203    329       PLASTOCYANIN-LIKE 2.
FT   DOMAIN      348    684       F5/8 TYPE A 2.
FT   DOMAIN      348    526       PLASTOCYANIN-LIKE 3.
FT   DOMAIN      536    684       PLASTOCYANIN-LIKE 4.
FT   DOMAIN      692   1573       B.
FT   DOMAIN      895    928       2 X 17 AA TANDEM REPEATS.
FT   REPEAT      895    911       1-1.
FT   REPEAT      912    928       1-2.
FT   SIMILAR    1135   1148       TO 14 AA REPEATS IN BOVINE FA5.
FT   DOMAIN     1185   1501       35 X 9 AA APPROXIMATE TANDEM REPEATS OF
FT                                [TNP]-L-S-P-D-L-S-Q-T.
FT   REPEAT     1185   1193       2-1.
FT   REPEAT     1194   1202       2-2.
FT   REPEAT     1203   1211       2-3.
FT   REPEAT     1212   1220       2-4.
FT   REPEAT     1221   1229       2-5.
FT   REPEAT     1230   1238       2-6.
FT   REPEAT     1239   1247       2-7.
FT   REPEAT     1248   1256       2-8.
FT   REPEAT     1257   1265       2-9.
FT   REPEAT     1266   1274       2-10.
FT   REPEAT     1275   1283       2-11.
FT   REPEAT     1284   1292       2-12.
FT   REPEAT     1293   1301       2-13.
FT   REPEAT     1302   1310       2-14.

  Query Match             6.2%;  Score 93;  DB 1;  Length 2224;
  Best Local Similarity   19.2%;  Pred. No. 17;
  Matches   58;  Conservative   42;  Mismatches  124;  Indels   78;  Gaps   11;

Qy          4 SNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIR 63
              ::::|  |  |::   ||:     ::   | |  : ||         |:: |     :|
Db        126 ASYLDHTFPAEKM--DDAVAPGREYTYEWSISEDSGPT---------HDDPP--CLTHIY 172

Qy         64 RQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRD 123
                |    |    ||  |:          | :    |    ||: : |  | |  :
Db        173 YSHENLIEDFNSGLIGPLLICKK---------GTLTEGGTQKTFDKQIVLLFAVFDESK- 222

Qy        124 ESGIAFVGLDIEWRPSFR-----KGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHL 178
                      |  |        | : | :  : :|          |  |  |  :
Db        223 -----------SWSQSSSLMYTVNGYVNGTMPDITVCA-----------HDHISWHLLGM 260

Qy        179 IEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLT------ 232
                |    |  :|  ::  |    :|  |   |  ||  :| : || ::|||
Db        261 SSGPEL--FSIHFNGQVLEQNHHKVSAITLVSATSTTANMTVGPEGKWIISSLTPKHLQA 318

Qy        233 ------------ETLVCKELLKPNRIRLGNWEFYPLSKQQL-QYAAT------DAYASW 272
                         :|   |:: :  |  :  ||::  ::: :  ||          | |
Db        319 GMQAYIDIKNCPKKTRNLKKITREQRRHMKRWEYFIAAEEVIWDYAPVIPANMDKKYRSQ 378

Qy        273 HL 274
              ||
Db        379 HL 380
```

```
RESULT 7
IRA2_YEAST
ID   IRA2_YEAST     STANDARD;      PRT;  3079 AA.
AC   P19158; Q08239;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Inhibitory regulator protein IRA2.
GN   IRA2 OR GLC4 OR CCS1 OR YOL081W OR O0985.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90318397; PubMed=2164637;
RA   Tanaka K., Nakafuku M., Tamanoi F., Kaziro Y., Matsumoto K., Toh-E A.;
RT   "IRA2, a second gene of Saccharomyces cerevisiae that encodes a
RT   protein with a domain homologous to mammalian ras GTPase-activating
RT   protein.";
RL   Mol. Cell. Biol. 10:4303-4313(1990).
RN   [2]
RP   SEQUENCE OF 1-2423 FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=95208358; PubMed=7900427;
RA   Zumstein E., Griffin H., Schweizer M.;
RT   "Sequence of a 10.27 kb segment on the left arm of chromosome XV from
RT   Saccharomyces cerevisiae includes part of the IRA2 gene and a
RT   putative new gene.";
RL   Yeast 10:1383-1387(1994).
RN   [3]
RP   SEQUENCE OF 1982-3079 FROM N.A.
RX   MEDLINE=97321807; PubMed=9178509;
RA   Tzermia M., Katsoulou C., Alexandraki D.;
RT   "Sequence analysis of a 33.2 kb segment from the left arm of yeast
RT   chromosome XV reveals eight known genes and ten new open reading
RT   frames including homologues of ABC transporters, inositol
RT   phosphatases and human expressed sequence tags.";
RL   Yeast 13:583-589(1997).
RN   [4]
RP   IDENTIFICATION OF CCS1 AS IRA2.
RX   MEDLINE=92405229; PubMed=1326414;
RA   Bussereau F., Dupont C.H., Boy-Marcotte E., Mallet L., Jacquet M.;
RT   "The CCS1 gene from Saccharomyces cerevisiae which is involved in
RT   mitochondrial functions is identified as IRA2 an attenuator of RAS1
RT   and RAS2 gene products.";
RL   Curr. Genet. 21:325-329(1992).
CC   -!- FUNCTION: INHIBITORY REGULATOR OF THE RAS-CYCLIC AMP PATHWAY.
CC       STIMULATES THE GTPASE ACTIVITY OF RAS PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 RAS-GAP DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33779; AAA34710.1; -.
DR   EMBL; X83121; CAA58201.1; -.
DR   EMBL; X75449; CAA53202.1; -.
DR   EMBL; Z74823; CAA99093.1; -.
DR   PIR; S11190; RGBYI2.
DR   SGD; S0005441; IRA2.
DR   InterPro; IPR001936; RasGAP.
DR   Pfam; PF00616; RasGAP; 1.
DR   SMART; SM00323; RasGAP; 1.
DR   PROSITE; PS00509; RAS_GTPASE_ACTIV_1; 1.
DR   PROSITE; PS50018; RAS_GTPASE_ACTIV_2; 1.
KW   GTPase activation.
FT   DOMAIN      1701   1890       RAS-GAP.
FT   DOMAIN       399    409       POLY-SER.
FT   DOMAIN       412    416       POLY-SER.
FT   DOMAIN       520    528       POLY-ALA.
FT   DOMAIN      2469   2472       POLY-LEU.
FT   CONFLICT    2317   2317       I -> K (IN REF. 3).
SQ   SEQUENCE   3079 AA;  351631 MW;  651EB2A2EEB479C0 CRC64;

  Query Match             6.2%;  Score 93;  DB 1;  Length 3079;
  Best Local Similarity   19.3%;  Pred. No. 26;
  Matches   57;  Conservative   43;  Mismatches  112;  Indels   84;  Gaps   10;

Qy          9 DAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPR 68
              | | :   | || :       :   |:       :    :  | :|  ::   : |   |
Db       1976 DDFYKTTFLLIDDVLGQLGQPKMEFSN-------EIPIYIREHMDDYPELYEFMNRHAFR 2028

Qy         69 SITSSTSY----------KRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIK-- 116
              :| :||:|          :  |:      ||           |    ::|  | : ::
Db       2029 NIETSTAYSPSVHESTSSEGIPIITLTMSNF----------SDRHVDIDTVAYKFLQIY 2077

Qy        117 ----------VLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHI 166
                       ::|    : |    |||:    |   |:|| :||  : |:   | :|
Db       2078 ARIWTTKHCLIIDCTEFDEG----GLDMRKFISLVMGLLP-EVAP-KNCIGCYYFNVNET 2131

Qy        167 F---------------HSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 211
              |               | ||    :   |  |:|  :|| |  :|:  |  ||: |:
Db       2132 FMDNYGKCLDKDNVYVSSKIPHYFINSNSDEGLMK-SVGITGQGLKVLQDIRVSLHDITL 2190

Qy        212 LSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATD 267
                :  |:                          |  :::|:  |  |:  ||   |
Db       2191 YDEKRNR---------------------FTPVSLKIGDIYFQVLHETPRQYKIRD 2224


RESULT 8
LEU2_SCHPO
ID   LEU2_SCHPO     STANDARD;      PRT;   758 AA.
AC   O14289;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   3-isopropylmalate dehydratase (EC 4.2.1.33) (Isopropylmalate
DE   isomerase) (Alpha-IPM isomerase) (IPMI).
GN   SPAC9E9.03.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
```

```
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate
CC       and 3-isopropylmalate, via the formation of 2-isopropylmaleate.
CC   -!- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmaleate +
CC       H(2)O.
CC   -!- CATALYTIC ACTIVITY: 2-isopropylmaleate + H(2)O = 2-
CC       isopropylmalate.
CC   -!- PATHWAY: Leucine biosynthesis; second step.
CC   -!- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z99262; CAB16402.1; -.
DR   InterPro; IPR000573; Aconitase_C.
DR   InterPro; IPR001030; Aconitase_N.
DR   InterPro; IPR004430; LeuC.
DR   InterPro; IPR004431; LeuD.
DR   Pfam; PF00330; aconitase; 1.
DR   Pfam; PF00694; Aconitase_C; 1.
DR   PRINTS; PR00415; ACONITASE.
DR   ProDom; PD000511; Aconitase_N; 1.
DR   TIGRFAMs; TIGR00170; leuC; 1.
DR   TIGRFAMs; TIGR00171; leuD; 1.
DR   PROSITE; PS00450; ACONITASE_1; 1.
DR   PROSITE; PS01244; ACONITASE_2; 1.
KW   Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S.
FT   METAL       359    359       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT   METAL       420    420       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT   METAL       423    423       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
SQ   SEQUENCE   758 AA;  82782 MW;  CE78C36828380E47 CRC64;

  Query Match             6.2%;  Score 92;  DB 1;  Length 758;
  Best Local Similarity   18.2%;  Pred. No. 4.9;
  Matches   61;  Conservative   53;  Mismatches  107;  Indels  114;  Gaps   17;

Qy         34 SSSSAAPTVQATTS------VHGHEEDPNQIPNNIR-RQLPRS-----ITSSTSYKRFPL 81
              :::   |||   ||       ::|:  ||  : :|:|   : ||     :  :||   :|:
Db        292 NAADVLPTVTWGTSPQDVIPINGNIPDPAHVKDNVRAASIQRSLEYMGLKPNTSIVSYPI 351

Qy         82 SR-----CRARNFPAMRFGGRILYSKTATEVDKRAM-----QLIKVLDTKRDESGI--AF 129
               :     |       :|    ::  :      | ||      |:| :    :  |:   |
Db        352 DKVFIGSCTNSRIEDLRLAAAVVKGRKVAANVKDAMIVPGSGLVKKM---AEAEGLDQIF 408

Qy        130 VGLDIEWRP---SFRKGVLPGKVATVQIC----------------------------V 156
              :    :||    |   |: | ::   : |                            :
Db        409 IEAGFDWREAGCSMCLGMNPDQLKPYERCASTSNRNFEGRQGAKGRTHLVSPAMAAAAAI 468

Qy        157 DSNYCDVMHIF---HSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDL- 212
                : |:|   |    :| |  : :   | :       ::|       | |:|: |  |
Db        469 KGHLCNVREFFGDVSNGSPSIITNKNYDPS-----HDVEG-------DIGLSVDDATDAV 516

Qy        213 --SDLANQKIGGDKKWGLASL------------------TETLVCKELLKP-NRIRLGN 250
                :|     : |    | | :                    |: :: |: ||   |  ||
Db        517 TDADGIATNVAGSVSSGSAGIPKFTVVEGIAAPLPMANVDTDKIIPKQFLKTIKRTGLGQ 576

Qy        251 WEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAV 285
              : ||     :::| |             |::|| |
Db        577 FAFY-----EIRYDADG----------KEIPDFV 595


RESULT 9
VL2_HPV03
ID   VL2_HPV03      STANDARD;      PRT;   473 AA.
AC   P36744;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Minor capsid protein L2.
GN   L2.
OS   Human papillomavirus type 3.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10614;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94265501; PubMed=8205838;
RA   Delius H., Hofmann B.;
RT   "Primer-directed sequencing of human papillomavirus types.";
RL   Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74462; CAA52473.1; -.
DR   PIR; S36553; S36553.
DR   InterPro; IPR000784; Late_L2.
DR   Pfam; PF00513; late_protein_L2; 1.
KW   Coat protein; Late protein.
SQ   SEQUENCE   473 AA;  50662 MW;  DD8298FE517A04DD CRC64;

  Query Match             6.1%;  Score 91.5;  DB 1;  Length 473;
  Best Local Similarity   21.6%;  Pred. No. 2.9;
  Matches   66;  Conservative   34;  Mismatches   95;  Indels  111;  Gaps   14;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              :||:|: : ||||  || :           :   |        |:  ||:||
Db        153 VSSTNFSNPAFTEPSLLEVP----------QNGEVSGHILISTPTSGTHGYEE------- 195

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAM-RFGGRILYSKTATEVDKRAMQLIKVLD 119
                :|   :|     |:|   :   | : |  |  |||| |:|       || |
Db        196 -----IPMETFASPGTGTEPIS---STPVPGVSRIAGPRLYSKAVTQV--------KVTD 239

Qy        120 ---TKRDESGIAFVGLDIEWRPS---FRKGVLPGKVATVQICVDSNYCDVMHIF------ 167
                 |  | : |      |       | :   | :|       ||:: |:: :
Db        240 PAFLTRPRSLMTFDNPVFEPEDETIIFERPYSPSQVP------DSDFLDILRLHRPALTS 293

Qy        168 ------HSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIG 221
                   :| : | |       |   : | |:            |  :   :||| :
Db        294 RRGTVRYSRVGQKL------SMRTRSGKGL----------GARVHYYQDLSPIGP---- 332

Qy        222 GDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDL 281
                    || :  : |: |                    |:  || |  || |  |:
Db        333 ----------TEDIEMEPLIAP------------------ASASAYDS--LYDVYADV 360

Qy        282 PDAVSG 287
               ||  |
Db        361 DDADIG 366


RESULT 10
CPV1_MOUSE
ID   CPV1_MOUSE     STANDARD;      PRT;   503 AA.
AC   P28649;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Cytochrome P450 19 (Aromatase) (EC 1.14.14.1) (CYPXIX) (Estrogen
DE   synthetase) (P-450AROM).
GN   CYP19 OR AROM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=91378248; PubMed=1897929;
RA   Terashima M., Toda K., Kawamoto T., Kuribayashi I., Ogawa Y.,
RA   Maeda T., Shizuta Y.;
RT   "Isolation of a full-length cDNA encoding mouse aromatase P450.";
RL   Arch. Biochem. Biophys. 285:231-237(1991).
CC   -!- FUNCTION: CATALYZES THE FORMATION OF AROMATIC C18 ESTROGENS FROM
CC       C19 ANDROGENS.
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00659; BAA00551.1; -.
DR   PIR; S13912; S13912.
DR   MGD; MGI:88587; Cyp19.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Electron transport; Oxidoreductase; Monooxygenase; Membrane;
KW   Heme.
FT   BINDING     437    437       HEME (BY SIMILARITY).
SQ   SEQUENCE   503 AA;  58015 MW;  737400389D1AEFF1 CRC64;

  Query Match             6.1%;  Score 91;  DB 1;  Length 503;
  Best Local Similarity   23.2%;  Pred. No. 3.5;
  Matches   38;  Conservative   23;  Mismatches   33;  Indels   70;  Gaps    8;

Qy         93 RFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEW---RPSFRKGVL-PGK 148
              |||             || :| | :      |:|| |      |   || | | :  ||
Db        115 RFG-----------SKRGLQCIGM-----HENGIIFNNNPSLWRTIRPFFMKALTGPGL 157

Qy        149 VATVQICVD--------------SNYCDVMHIFHSGIPQSLQHLIED-STLVKVGIGID 192
              |  |::||:              | | ||: :        ::|:: | | :: :|| :|
Db        158 VRMVEVCVESIKQHLDRLGEVTDTSGYVDVLTL--------MRHIMLDTSNMLFLGIPLD 209

Qy        193 GDSVK-----------------------LFHDYGVSIKDVED 211
                ::                      |:  |  |:||::|
Db        210 ESAIVKKIQGYFNAWQALLIKPNIFFKISWLYRKYERSVKDLKD 253


RESULT 11
DPO1_HAEIN
ID   DPO1_HAEIN     STANDARD;      PRT;   930 AA.
AC   P43741;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   DNA polymerase I (EC 2.7.7.7) (POL I).
GN   POLA OR HI0856.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC       EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + [DNA](N).
CC   -!- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
CC   -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32767; AAC22515.1; -.
DR   HSSP; P00582; 1KFS.
DR   TIGR; HI0856; -.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   InterPro; IPR002421; 5_3_exonuclease.
DR   InterPro; IPR001098; DNA_pol.
DR   InterPro; IPR002298; DNA_polI.
DR   InterPro; IPR000513; Exo_N_I.
DR   InterPro; IPR003584; HHH_2.
DR   Pfam; PF00476; DNA_pol_A; 1.
DR   Pfam; PF01367; 5_3_exonuclease; 1.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   Pfam; PF02739; 5_3_exonuc_N; 1.
DR   PRINTS; PR00868; DNAPOLI.
DR   SMART; SM00474; 35EXOc; 1.
DR   SMART; SM00475; 53EXOc; 1.
DR   SMART; SM00279; HhH2; 1.
DR   SMART; SM00482; POLAc; 1.
DR   TIGRFAMs; TIGR00593; pola; 1.
DR   PROSITE; PS00447; DNA_POLYMERASE_A; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW   Hydrolase; Exonuclease; DNA-binding; Complete proteome.
SQ   SEQUENCE   930 AA;  103740 MW;  226654BB7CFF730B CRC64;

  Query Match             6.0%;  Score 89.5;  DB 1;  Length 930;
  Best Local Similarity   20.7%;  Pred. No. 11;
  Matches   61;  Conservative   42;  Mismatches   93;  Indels   99;  Gaps   15;

Qy         24 ASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPRSITSSTSYKRFPLSR 83
              | | | |  :   : | ::  ||      | |                    ||     :
Db        280 ARYEFKRWLNEVMNGADSITQTT------EQP--------------VKMNQYKATSQDQ 318

Qy         84 CRARNFPAMRFGGRILYSKTATEVD-------KRAMQLIKVLDTKRDE--------SGIA 128
                | | ::   |  |    |: |          | :|| | ||: |           ||:
Db        319 SAVENTPKIQI-DRTKYETLLTQADLTRWIEKLNAAKLIAV-DTETDSLDYMSANLVGIS 376

Qy        129 FVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQH---------LI 179
              |              :  |: | : : :|  | |         |::|:          ::
Db        377 F------------ALENGEAAYLPLQLD--YLDA--------PKTLEKSTALAAIKPIL 413

Qy        180 EDSTLVKVGIGIDGDSVKLFHDYGVSIKDVE---------------DLSDLANQKIGGD 223
              |:  : |:|  |  |   :|  :|: :: ||                :: ||| : :|
Db        414 ENPNIHKIGQNIKFDE-SIFARHGIELQGVEFDTMLLSYTLNSTGRHNMDDLAKRYLG-- 470

Qy        224 KKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVL 278
                    ||:  : |    : :|  :   || :|  :||| ||  :  | : |
Db        471 -------HETIAFESLAGKGKSQL-TFNQIPL-EQATEYAAEDADVTMKLQQAL 515
```

```
RESULT 12
MSH6_ARATH
ID   MSH6_ARATH      STANDARD;      PRT;  1324 AA.
AC   O04716;
DT   15-DEC-1998 (Rel. 37, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA mismatch repair protein MSH6-1 (AtMsh6-1).
GN   MSH6-1 OR AGAA.3 OR AT4G02070 OR T10M13.8.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Landsberg erecta;
RA   Till S., Granat S., Parnell L., Kaplan N., Hoffman J., Lodhi M.,
RA   Johnson A.F., Dedhia N., Martienssen R., McCombie W.R.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=99444907; PubMed=10517319;
RA   Ade J., Belzile F., Philippe H., Doutriaux M.P.;
RT   "Four mismatch repair paralogues coexist in Arabidopsis thaliana:
RT   AtMSH2, AtMSH3, AtMSH6-1 and AtMSH6-2.";
RL   Mol. Gen. Genet. 262:239-249(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
CC   -!- FUNCTION: INVOLVED IN POST-REPLICATIVE DNA-MISMATCH REPAIR (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: HETERODIMER OF MSH2 AND MSH6 (GTBP) (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF001535; AAB57798.1; -.
DR   EMBL; AJ245967; CAB53337.1; -.
DR   EMBL; AF001308; AAC78699.1; -.
DR   EMBL; AL161493; CAB80700.1; -.
DR   InterPro; IPR000432; MutS_C.
DR   InterPro; IPR002863; MutS_N.
DR   InterPro; IPR002999; Tudor.
DR   Pfam; PF00488; MutS_C; 1.
DR   Pfam; PF01624; MutS_N; 2.
DR   ProDom; PD001263; MutS_C; 1.
DR   SMART; SM00534; MUTSac; 1.
DR   SMART; SM00533; MUTSd; 1.
DR   SMART; SM00333; TUDOR; 1.
DR   PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW   DNA repair; ATP-binding; DNA-binding.
FT   NP_BIND    1083   1090       ATP (POTENTIAL).
FT   CONFLICT    316    318       ADR -> GKPKSLLL (IN REF. 1).
FT   CONFLICT    383    383       Q -> QVRRAHGNL (IN REF. 1).
FT   CONFLICT    852    852       P -> PGIINLLHIHILILANCTASHIISLP (IN
FT                                REF. 1).
SQ   SEQUENCE   1324 AA;  146797 MW;  787A340272CF979C CRC64;

  Query Match             6.0%;  Score 89.5;  DB 1;  Length 1324;
  Best Local Similarity   23.2%;  Pred. No. 17;
  Matches   68;  Conservative   40;  Mismatches   92;  Indels   93;  Gaps   18;

Qy         66 LPRSITSSTSYKRFP-LSRCRARNFPAMRFGGR-----ILYSKTATEVDKRAMQLIKVL- 118
              || |:    |  | | : |  || | ::   ||     :|| ||    |:  : |  |
Db        766 LPYSLEFRKSLSRLPDMERLIARMFSSIEASGRNGDKVVLYEDTA---KKQVQEFISTLR 822

Qy        119 ------------------DTKRD-------------ESGIAFVGLDIEWRPSFRKG-VL 145
                               || |                | | :     :|  :   | |:
Db        823 GCETMAEACSSLRAILKHDTSRRLLHLLTPGQSLPNISSSIKYFKDAFDWVEAHNSGRVI 882

Qy        146 PGKVATVQICVDSNY---CDVMHIFHSGIPQSLQH---LIEDSTLVKVGIGID------- 192
              | : |      |  |   |  :  | | : : |:    |: |:::  | :| |
Db        883 PHEGA------DEEYDCACKTVEEFESSLKKHLKEQRKLLGDASINYVTVGKDEYLLEVP 936

Qy        193 -GDSVKLFHDY-------GVS------IKDVEDLSDLANQKIGGDKKWGLASLTETLV-- 236
                | : |||       |||      || :  | :|:  |   :|:  | |::: |:
Db        937 ESLSGSVPHDYELCSSKKGVSRYWTPTIKKL--LKELSQAK--SEKESALKSISQRLIGR 992

Qy        237 -CKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
               |:   |        |     :  :|    : |:||   |: ::  |  :|||
Db        993 FCEHQEK--------WRQLVSATAELDVLISLAFAS-DSYEGVRCRP-VISGS 1035


RESULT 13
BLM_DROME
ID   BLM_DROME      STANDARD;      PRT;  1487 AA.
AC   Q9VGI8; Q9Y062;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
```

```
DE   Bloom's syndrome protein homolog (EC 3.6.1.-) (Dmblm) (Mutagen-
DE   sensitive protein 309) (RecQ helicase homolog).
GN   mus309 or blm or CG6920.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., AND REPEATS.
RC   STRAIN=Canton-S;
RX   MEDLINE=99160561; PubMed=10049920;
RA   Kusano K., Berres M.E., Engels W.R.;
RT   "Evolution of the RECQ family of helicases: a Drosophila homolog,
RT   Dmblm, is similar to the human Bloom syndrome gene.";
RL   Genetics 151:1027-1039(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: Participates in DNA replication and may participate in
CC       repair. Exhibits a magnesium dependent ATP-dependent DNA-helicase
CC       activity that unwinds single- and double-stranded DNA in a 3'-5'
CC       direction (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECQ SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 HRDC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U92536; AAD41441.1; -.
DR   EMBL; AE003692; AAF54691.1; -.
DR   FlyBase; FBgn0002906; mus309.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002464; DEAH_box.
DR   InterPro; IPR002121; HRDC.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004589; RecQ.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00570; HRDC; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00341; HRDC; 1.
DR   TIGRFAMs; TIGR00614; recq; 1.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; 1.
KW   Hydrolase; Helicase; ATP-binding; DNA-binding; DNA replication;
KW   Repeat; Nuclear protein.
FT   DOMAIN       89    138       2 X 24AA REPEATS OF L-D-L-S-V-S-P-L-A-E-
FT                                L-[SP]-A-K-K-K-[YS]-[AD]-R-D-[SP]-P-P-K-
FT                                P.
FT   REPEAT       89    112       1.
FT   REPEAT      115    138       2.
FT   DOMAIN      225    230       POLY-PRO.
FT   DOMAIN      464    467       POLY-SER.
FT   DOMAIN     1283   1363       HRDC.
FT   DOMAIN     1369   1372       POLY-GLU.
FT   DOMAIN     1416   1432       NUCLEAR LOCALIZATION SIGNAL (BY
FT                                SIMILARITY).
FT   NP_BIND     759    766       ATP (BY SIMILARITY).
FT   SITE        865    868       DEAH BOX.
FT   CONFLICT     98     98       K -> E (IN REF. 1).
FT   CONFLICT    110    110       K -> P (IN REF. 1).
FT   CONFLICT    126    126       L -> P (IN REF. 1).
FT   CONFLICT    134    136       SPK -> PPP (IN REF. 1).
FT   CONFLICT    169    169       Q -> P (IN REF. 1).
FT   CONFLICT    222    222       P -> S (IN REF. 1).
FT   CONFLICT    299    299       Y -> N (IN REF. 1).
FT   CONFLICT    417    417       M -> V (IN REF. 1).
FT   CONFLICT    459    459       Q -> R (IN REF. 1).
FT   CONFLICT    471    471       S -> C (IN REF. 1).
FT   CONFLICT    496    496       H -> P (IN REF. 1).
FT   CONFLICT    594    594       D -> G (IN REF. 1).
FT   CONFLICT    614    614       G -> A (IN REF. 1).
FT   CONFLICT    805    805       K -> E (IN REF. 1).
FT   CONFLICT   1116   1116       E -> G (IN REF. 1).
SQ   SEQUENCE   1487 AA;  166077 MW;  07361B8005E29432 CRC64;

  Query Match             6.0%;  Score 89.5;  DB 1;  Length 1487;
  Best Local Similarity   24.5%;  Pred. No. 20;
  Matches   64;  Conservative   42;  Mismatches   94;  Indels   61;  Gaps   16;

Qy         60 NNIR-RQLP-RSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKV 117
              :|:| | || : :::     |:  |:          | | |:|  :  | |: : :: |
Db        930 SNLRYRVLPKKGVSTLDDISRYIRSK-------PQHFSG-IIYCLSRKECDETSKKMCK- 980

Qy        118 LDTKRDESGIAFV----GLDIEWRPSFRKGVLPGKV----ATVQICVDSNYCDVMHIFHS 169
                        |:  |    ||    | | :|  | ||:    |||   :  :  ||  : |
Db        981 -------DGVRAVSYHAGLTDTDRESRQKDWLTGKMRVICATVAFGMGIDKPDVRFVLHY 1033

Qy        170 GIPQSLQHLIEDSTLVKVGIGIDGDSVK--LFHDYG--VSIKDVEDLSDLANQ----KIG 221
               :|:|::   :::       | |||     |:::|   : || : | || | |    ||
Db       1034 SLPKSIEGYYQEAGRA----GRDGDVADCILYYNYSDMLRIKKMLD-SDKALQYNVKKIH 1088

Qy        222 GDKKWGLASLTETLV-CKELLKPNRIRLGNWEFYPLSKQQLQ--------------YAAT 266
               |  : :    | |  |:        :|  :  :  |:| |:              | |
Db       1089 VDNLYRIVGYCENLTDCRRAQ-----QLDYFGEHFTSEQCLENRETACDNCINKRAYKAV 1143

Qy        267 DAYASWHLYKVLKDLPDAVSG 287
               ||    |  |  : : |  ||
Db       1144 DALE--HARKAARAVKDLCSG 1162
```

```
RESULT 14
NK2B_BRARE
ID   NK2B_BRARE     STANDARD;      PRT;   269 AA.
AC   Q90481;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Homeobox protein NK-2 homolog B (Homeobox protein NKX-2.2).
GN   NKX2-2 OR NKX2.2 OR NK2.2.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95324401; PubMed=7600991;
RA   Barth K.A., Wilson S.W.;
RT   "Expression of zebrafish nk2.2 is influenced by sonic
RT   hedgehog/vertebrate hedgehog-1 and demarcates a zone of neuronal
RT   differentiation in the embryonic forebrain.";
RL   Development 121:1755-1768(1995).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A CONTINUOUS NARROW BAND OF CELLS
CC       ALONG A BOUNDARY ZONE DEMARCATING THE LOCATION AT WHICH TWO OF THE
CC       EARLIEST NUCLEI IN THE BRAIN DIFFERENTIATE.
CC   -!- SIMILARITY: BELONGS TO THE NK-2 HOMEOBOX FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 HOMEOBOX DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X85977; CAA59967.1; -.
DR   HSSP; P22808; 1NK3.
DR   ZFIN; ZDB-GENE-980526-403; nkx2.2.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND    125    184       HOMEOBOX.
SQ   SEQUENCE   269 AA;  30306 MW;  CF006285CDD33D63 CRC64;

  Query Match             5.9%;  Score 88;  DB 1;  Length 269;
  Best Local Similarity   22.2%;  Pred. No. 2.7;
  Matches   45;  Conservative   32;  Mismatches   64;  Indels   62;  Gaps    8;

Qy          8 DDAFTEEELLAIDAIEAS-YNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQL 66
              |: :|   |   |:|: | :  | :|  :|: :|           |   :: |:| :
Db         69 DNPYT-RWLATTDSIQYSLHGLSANSQDTSAKSP----------EPSADESPDNDKETS 116

Qy         67 PRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTAT-EVDKRAMQ----------- 113
                 |    ||                 |:|:||  | |:::|  |
Db        117 SNGSDSGKKRKR----------------RVLFSKAQTYELERRFRQQRYLSAPEREHL 158

Qy        114 --LIKVLDTK----------RDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYC 161
                ||::  |:         : :   |  |:::   || |:  :|  |   : |
Db        159 ASLIRLTPTQVKIWFQNHRYKMKRARAEKGMEVTHLPSPRRVAVPVLVRDGKPCHTLKAQ 218

Qy        162 DVMHIFHSGIP------QSLQHL 178
              |:   | :|||      |||||:
Db        219 DLAATFQAGIPFSAYSAQSLQHM 241
```

```
RESULT 15
PROA_VIBCH
ID   PROA_VIBCH     STANDARD;      PRT;   416 AA.
AC   Q9KPT9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Gamma-glutamyl phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-
DE   semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde
DE   dehydrogenase) (GSA dehydrogenase).
GN   PROA OR VC2273.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- FUNCTION: CATALYZES THE NADPH DEPENDENT REDUCTION OF L-GAMMA-
CC       GLUTAMYL 5-PHOSPHATE INTO L-GLUTAMATE 5-SEMIALDEHYDE AND
CC       PHOSPHATE. THE PRODUCT SPONTANEOUSLY UNDERGOES CYCLIZATION TO FORM
CC       1-PYRROLINE-5-CARBOXYLATE.
CC   -!- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate +
CC       NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
CC   -!- PATHWAY: Proline biosynthesis; second step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004299; AAF95417.1; ALT_INIT.
DR   TIGR; VC2273; -.
DR   InterPro; IPR002086; Aldehyde_dehydr.
DR   InterPro; IPR000965; Gglut_pp_reduct.
DR   Pfam; PF00171; aldedh; 1.
DR   TIGRFAMs; TIGR00407; proA; 1.
DR   PROSITE; PS01223; PROA; FALSE_NEG.
KW   Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
SQ   SEQUENCE   416 AA;  44504 MW;  7ACA58E537235563 CRC64;

  Query Match             5.9%;  Score 88;  DB 1;  Length 416;
  Best Local Similarity   21.7%;  Pred. No. 4.8;
  Matches   69;  Conservative   54;  Mismatches  109;  Indels   86;  Gaps   19;

Qy          7 IDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNI---R 63
              : ||  :  ||         | ||  : ::     ::    | | | |   : | :   |
Db         59 LSDAMLDRLLL---------NESRLQAIANDVRNVIKLNDPV-GSEIDSRVLENGMSLAR 108

Qy         64 RQLPRSITSSTSYKRFP------LSRC-RARNFPAMRFGGRILYSKTATEVDKRAMQLIK 116
              |::|  :     |:  |       : | :  |   :| |    :|          |:|:|
Db        109 RRVPLGVV-GVIYEARPNVTIDIAALCLKTGNAAILRGGKETFFSN---------MELVK 158

Qy        117 VLDTKRDESGIAFVGLD-IEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQ-- 173
              |: :  |::|:    :  ||         | :    |:     :| |::      ||:
Db        159 VIQSALDKAGLPAASVQYIE---------KPDRELVTQLLKMDDYVDMI------IPRGG 203
```

```
Qy        174 -SLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDL-------ANQKIGGDKK 225
                 |  : :::: | | ||           :|:|   |:: :|||        | |:   :
Db        204 AGLHKMCKENSTVPVIIG----------GFGISHIFVDESADLDKSVAVIENAKV--QRP 251

Qy        226 WGLASLTETLVCKELLKPNRIRL-----GNWEFY--PLSKQQLQYAA--TDAYA-----S 271
                 :|   || : : ||   :|     |   |   | :|  :  ||   || |
Db        252 SACNALDTLLVHQAIAKPLLDKLIAKLNGKVAFVAEPKAKALMSSAAELRDAQAGDFDTE 311

Qy        272 WHLY----KVLKDLPDAV 285
                |  |    ||::|: :|:
Db        312 WLSYTLGVKVVQDVQEAI 329
```

Search completed: June 6, 2003, 10:22:52
Job time : 12 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 6, 2003, 10:19:55 ; Search time 32 Seconds
(without alignments)
1854.424 Million cell updates/sec

Title: US-09-896-186B-24
Perfect score: 1491
Sequence: 1 MSSSNWIDDAFTEEELLAID.....YASWHLYKVLKDLPDAVSGS 288

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1466.5	98.4	285	10	Q9FT68	Q9ft68 arabidopsis
2	1439.5	96.5	313	10	Q9SVM6	Q9svm6 arabidopsis
3	339.5	22.8	1436	13	O93530	O93530 xenopus lae
4	339	22.7	988	11	Q9Z241	Q9z241 mus musculu
5	339	22.7	1401	11	Q9JKD4	Q9jkd4 mus musculu
6	330.5	22.2	643	11	O35948	O35948 mus musculu
7	238	16.0	201	10	Q93VU9	Q93vu9 oryza sativ
8	229.5	15.4	346	5	Q9VE86	Q9ve86 drosophila
9	227.5	15.3	501	10	Q8S2J5	Q8s2j5 oryza sativ
10	221.5	14.9	353	5	Q961E1	Q961e1 drosophila
11	212.5	14.3	583	5	Q9VGN7	Q9vgn7 drosophila
12	179.5	12.0	582	10	Q9C7K6	Q9c7k6 arabidopsis
13	174.5	11.7	391	10	Q8SA18	Q8sa18 oryza sativ
14	173.5	11.6	494	10	Q93VS2	Q93vs2 oryza sativ
15	159.5	10.7	699	3	Q9C2I6	Q9c2i6 neurospora
16	158	10.6	123	10	O22964	O22964 arabidopsis
17	152.5	10.2	496	11	Q8VEG4	Q8veg4 mus musculu
18	151.5	10.2	199	10	Q9LK79	Q9lk79 arabidopsis
19	151	10.1	496	4	Q9NVH0	Q9nvh0 homo sapien
20	142	9.5	599	4	Q96NP1	Q96np1 homo sapien
21	127.5	8.6	239	10	Q9SIH3	Q9sih3 arabidopsis
22	121	8.1	505	10	Q9FIN8	Q9fin8 arabidopsis
23	112.5	7.5	217	10	O80886	O80886 arabidopsis
24	111.5	7.5	210	10	Q9LHG3	Q9lhg3 arabidopsis
25	111	7.4	242	10	Q9LHG2	Q9lhg2 arabidopsis
26	111	7.4	300	2	Q9AL74	Q9al74 citrobacter
27	110	7.4	238	10	Q9C7A5	Q9c7a5 arabidopsis
28	108	7.2	300	2	Q8RNV0	Q8rnv0 citrobacter
29	105	7.0	625	5	Q9VIF1	Q9vif1 drosophila
30	103	6.9	220	10	Q9LHG1	Q9lhg1 arabidopsis
31	102.5	6.9	710	5	Q9NA69	Q9na69 caenorhabdi
32	101	6.8	876	5	Q17951	Q17951 caenorhabdi
33	99.5	6.7	393	16	Q8R7C1	Q8r7c1 thermoanaer
34	98.5	6.6	265	10	Q9LHG5	Q9lhg5 arabidopsis
35	98.5	6.6	532	17	Q976Y3	Q976y3 sulfolobus
36	98	6.6	216	10	Q8SA46	Q8sa46 hordeum vul
37	97.5	6.5	445	17	Q9YDV0	Q9ydv0 aeropyrum p
38	97	6.5	579	2	Q9RQK2	Q9rqk2 bacillus st
39	97	6.5	906	5	Q17281	Q17281 botryllus s
40	96.5	6.5	488	16	Q97JZ0	Q97jz0 clostridium
41	95	6.4	455	10	Q9ZWS2	Q9zws2 vigna mungo
42	95	6.4	925	5	Q9GUG1	Q9gug1 caenorhabdi
43	95	6.4	4589	5	O76506	O76506 tetrahymena
44	94.5	6.3	353	10	Q9LHG4	Q9lhg4 arabidopsis
45	94.5	6.3	416	5	O96144	O96144 plasmodium

ALIGNMENTS

RESULT 1
Q9FT68
ID Q9FT68 PRELIMINARY; PRT; 285 AA.
AC Q9FT68;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Exonuclease.
GN WRNEXO.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA; TISSUE=FLOWER;
RA Hartung F., Plchova H., Puchta H.;
RT "Molecular characterisation of RecQ homologues in Arabidopsis
RT thaliana.";
RL Nucleic Acids Res. 21:4275-4282(2000).
DR EMBL; AJ404476; CAC14871.1; -.
DR InterPro; IPR002562; 3_5_exonuclease.
DR Pfam; PF01612; 3_5_exonuclease; 1.
DR SMART; SM00474; 35EXOc; 1.
SQ SEQUENCE 285 AA; 31750 MW; A198CB93653E229B CRC64;

Query Match 98.4%; Score 1466.5; DB 10; Length 285;
Best Local Similarity 99.0%; Pred. No. 6.7e-121;
Matches 285; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

```
Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
```

```
Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
              ||||||||||||||||||||||||||||||||||   |||||||||
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLY---KDLPDAVSGS 285


RESULT 2
Q9SVM6
ID   Q9SVM6      PRELIMINARY;      PRT;   313 AA.
AC   Q9SVM6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Hypothetical 35.1 kDa protein.
GN   F18A5.260 OR AT4G13870.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Weber N., Grueninger D., Schmidheini T., Bancroft I.,
RA   Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Weber N., Grueninger D., Schmidheini T., Mewes H.W., Lemcke K.,
RA   Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL035528; CAB36851.1; -.
DR   EMBL; AL161537; CAB78429.1; -.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   SMART; SM00474; 35EXOc; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   313 AA;  35105 MW;  448D7ACC375D4D22 CRC64;

  Query Match            96.5%;  Score 1439.5;  DB 10;  Length 313;
  Best Local Similarity  95.3%;  Pred. No. 1.8e-118;
  Matches  281;  Conservative    1;  Mismatches    4;  Indels    9;  Gaps    1;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKV---------LKDLPDAVS 286
              ||||||||||||||||||||||||||||||||||||          | ||   :|
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVTTTKNHLLTLNDLEAKIS 295


RESULT 3
O93530
ID   O93530      PRELIMINARY;      PRT;  1436 AA.
AC   O93530;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Werner syndrome helicase homolog (Focus forming activity 1).
GN   FFA-1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A., AND FUNCTION.
RX   MEDLINE=98361165; PubMed=9697700;
RA   Yan H., Chen C.-Y., Kobayashi R., Newport J.;
RT   "Replication focus-forming activity 1 and the Werner syndrome gene
RT   product.";
RL   Nat. Genet. 19:375-378(1998).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=96032793; PubMed=7569932;
RA   Yan H., Newport J.;
RT   "FFA-1, a protein that promotes the formation of replication centers
RT   within nuclei.";
RL   Science 269:1883-1885(1995).
CC   -!- FUNCTION: ESSENTIAL FOR THE FORMATION OF DNA REPLICATION FOCAL
CC       CENTERS; STABLY ASSOCIATES WITH FOCI ELEMENTS GENERATING BINDING
CC       SITES FOR RP-A. EXHIBITS A MAGNESIUM DEPENDENT ATP-DEPENDENT DNA-
CC       HELICASE ACTIVITY. MAY BE INVOLVED IN THE CONTROL OF GENOMIC
CC       STABILITY.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE RECQ SUBFAMILY OF HELICASES.
CC   -!- SIMILARITY: CONTAINS 1 HRDC DOMAIN.
DR   EMBL; AF067418; AAC63512.1; -.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR002121; HRDC.
DR   InterPro; IPR004589; RecQ.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00570; HRDC; 1.
DR   SMART; SM00474; 35EXOc; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00341; HRDC; 1.
DR   TIGRFAMs; TIGR00614; recq; 1.
KW   Hydrolase; Helicase; ATP-binding; DNA-binding; Nuclear protein.
FT   DOMAIN      461    467       POLY-ASP.
FT   DOMAIN     1098   1117       HRDC.
FT   NP_BIND     512    519       ATP (BY SIMILARITY).
FT   SITE        609    612       DEAH BOX.
SQ   SEQUENCE   1436 AA;  161850 MW;  1BEAF05A25B4E230 CRC64;

  Query Match            22.8%;  Score 339.5;  DB 13;  Length 1436;
  Best Local Similarity  35.0%;  Pred. No. 8e-21;
  Matches   82;  Conservative   43;  Mismatches   84;  Indels   25;  Gaps    7;

Qy         61 NIRRQLPRSITSSTSYKRF---PLSRCRAR----NFPAMRFGGRILYSKTATEV-----D 108
              :::|:||  ::      |      | |:      | | |:| | |:||  : :      |
Db          3 SLQRKLPEWMSVKQQEDRIDDAKKSFCKKNILEDNLPFMKFNGSIVYSYESNDCSLLSED 62

Qy        109 KRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIF- 167
              |: |         |  :  :| |||| | : ||   |||| :|:||    | : ||
```

```
Db         63 IRSSLL---------EEDV--LGFDIEWPPVYTKG-KTGKVALIQVCVSEKKCYLFHISP 110

Qy        168 HSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVVSIKDVEDLSDLANQKIGGDKKWG 227
              :| |: |: |:|| :: |||:||:|| || || : :|   :||::||||:   :||
Db        111 MAGFPKGLKRLLEDESVRKVGVGIEGDQWKLMSDYELKLKGFIELSEMANQKLRCKEKWT 170

Qy        228 LASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDL 281
                | : |  ::| |    |  ||: : |:: |  ||||||||   :|| |: :
Db        171 FNGLIKHLFKEQLYKRKSYRCSNWDIFLLTEDQKLYAATDAYAGLLIYKKLEGM 224


RESULT 4
Q9Z241
ID   Q9Z241      PRELIMINARY;      PRT;   988 AA.
AC   Q9Z241;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Wrn protein.
GN   WRN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Paeper B.W., Gayle M., Brady W., Swartz A., Gillett L.A., Alisch R.S.,
RA   Mulligan J., Galas D., Fu Y.-H.;
RT   "Genomic structure of the human Werner's gene and cloning of the mouse
RT   homolog.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF091216; AAC72359.1; -.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004589; RecQ.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00474; 35EXOc; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   TIGRFAMs; TIGR00614; recq; 1.
KW   ATP-binding; Helicase.
SQ   SEQUENCE   988 AA;  111513 MW;  DBD4D77FA55D5B2D CRC64;

  Query Match             22.7%;  Score 339;  DB 11;  Length 988;
  Best Local Similarity   36.6%;  Pred. No. 5.3e-21;
  Matches   74;  Conservative   42;  Mismatches   72;  Indels   14;  Gaps    5;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              | | : | | |:||  |::       | : :  :  :  :  || |:|| | ::    ||
Db         37 NLPFLEFPGSIVYSYEASD----CSFLSEDISMRLSDGDV--VGFDMEWPPIYK----PG 86

Qy        148 K---VATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG 203
              |   || :|:|| | | : ||    : || |: |:|: :: | |:||:||  ||  |:
Db         87 KRSRVAVIQLCVSENKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRDFD 146

Qy        204 VSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQY 263
              | ::   :|:|:||:|:   : | |  | : :: |:||   ||  ||  :||:: |  |
Db        147 VKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLY 206

Qy        264 AATDAYASWHLYKVLKDLPDAV 285
              |||||||   :|: | :| | |
Db        207 AATDAYAGLIIYQKLGNLGDTV 228


RESULT 5
Q9JKD4
ID   Q9JKD4      PRELIMINARY;      PRT;  1401 AA.
AC   Q9JKD4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   WRN protein.
GN   WRN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20221574; PubMed=10757812;
RA   Lombard D.B., Beard C., Johnson B., Marciniak R.A., Dausman J.,
RA   Bronson R., Buhlmann J.E., Lipman R., Curry R., Sharpe A.,
RA   Jaenisch R., Guarente L.;
RT   "Mutations in the WRN Gene in Mice Accelerate Mortality in a p53-Null
RT   Background.";
RL   Mol. Cell. Biol. 20:3286-3291(2000).
DR   EMBL; AF241636; AAF64490.1; -.
DR   MGD; MGI:109635; Wrn.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR002121; HRDC.
DR   InterPro; IPR004589; RecQ.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00570; HRDC; 1.
DR   SMART; SM00474; 35EXOc; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00341; HRDC; 1.
DR   TIGRFAMs; TIGR00614; recq; 1.
KW   ATP-binding; Helicase.
SQ   SEQUENCE   1401 AA;  157300 MW;  6CBC330CD072C670 CRC64;

  Query Match             22.7%;  Score 339;  DB 11;  Length 1401;
  Best Local Similarity   36.6%;  Pred. No. 8.5e-21;
  Matches   74;  Conservative   42;  Mismatches   72;  Indels   14;  Gaps    5;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              | | : | | |:||  |::       | : :  :  :  :  || |:|| | ::    ||
Db         37 NLPFLEFPGSIVYSYEASD----CSFLSEDISMRLSDGDV--VGFDMEWPPIYK----PG 86

Qy        148 K---VATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG 203
              |   || :|:|| | | : ||    : || |: |:|: :: | |:||:||  ||  |:
Db         87 KRSRVAVIQLCVSENKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRDFD 146

Qy        204 VSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQY 263
              | ::   :|:|:||:|:   : | |  | : :: |:||   ||  ||  :||:: |  |
Db        147 VKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLY 206

Qy        264 AATDAYASWHLYKVLKDLPDAV 285
              |||||||   :|: | :| | |
Db        207 AATDAYAGLIIYQKLGNLGDTV 228


RESULT 6
O35948
ID   O35948      PRELIMINARY;      PRT;   643 AA.
AC   O35948;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Homolog of human Werners syndrome protein.
GN   WRN OR WRN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wu J., He J., Mountz J.D.;
RT   "mouse WRN.";
```

```
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U97045; AAB87366.1; -.
DR   MGD; MGI:109635; Wrn.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR004589; RecQ.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00474; 35EXOc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   TIGRFAMs; TIGR00614; recq; 1.
KW   ATP-binding; Helicase.
SQ   SEQUENCE   643 AA;  72821 MW;  F9125C234ABBB757 CRC64;

  Query Match             22.2%;  Score 330.5;  DB 11;  Length 643;
  Best Local Similarity   33.3%;  Pred. No. 1.7e-20;
  Matches   75;  Conservative   48;  Mismatches   83;  Indels   19;  Gaps    6;

Qy         61 NIRRQLP-----RSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLI 115
              :::|: |     :|   :|  |           | | : | | |:||  |::       |
Db          5 SLQRKFPEWMSMQSQRCATEEKACVQKNVLEDNLPFLEFPGSIVYSYEASD----CSFLS 60

Qy        116 KVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGK---VATVQICVDSNYCDVMHIFHSGI- 171
              : :  :  :  :  || |:|| | ::    |||   || :|:||  | | : ||    :
Db         61 EDISMRLSDGDV--VGFDMEWPPIYK----PGKRSRVAVIQLCVSENKCYLFHISSMSVF 114

Qy        172 PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASL 231
              || |: |:|: :: | |:||:||  ||  |: | ::   :|:|:||:|:   : | |  |
Db        115 PQGLKMLLENKSIKKAGVGIEGDQWKLLRDFDVKLESFVELTDVANEKLKCAETWSLNGL 174

Qy        232 TETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYK 276
               : :: |:|||   ||  ||  :||:: |  |||||||| : :|:
Db        175 VKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLYAATDAYAGFIIYR 219


RESULT 7
Q93VU9
ID   Q93VU9      PRELIMINARY;      PRT;   201 AA.
AC   Q93VU9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   P0487H02.22 protein (P0682B08.3 protein).
GN   P0487H02.22 OR P0682B08.3.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0487H02.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0682B08.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002883; BAB67862.1; -.
DR   EMBL; AP003578; BAB60933.1; -.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
SQ   SEQUENCE   201 AA;  21899 MW;  85F35C0CC5386265 CRC64;

  Query Match             16.0%;  Score 238;  DB 10;  Length 201;
  Best Local Similarity   31.0%;  Pred. No. 4.8e-13;

  Matches   58;  Conservative   36;  Mismatches   87;  Indels    6;  Gaps    4;

Qy         92 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVAT 151
              :|||   : :   ::|          |    |   |:  |||| ||:|:        |||
Db         10 VRFGSATIDTTVTSDVAAADEWARGVRAAARGGRGL-IVGLDCEWKPN-HVSWKTSKVAV 67

Qy        152 VQICVDSNYCDVMHIFHSG-IPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVE 210
              :|:|    :| |: :|::  :| ::  |: | ::  ||||:  |: ||  ||||
Db         68 LQLCAGERFCLVLQLFYANRVPPAVADLLGDPSVRLVGIGVGEDAAKLEADYGVWCAAPV 127

Qy        211 DLSDLANQK---IGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATD 267
              || |  |::   :|  :: ||    ::   : || |: : |||   |   |::||  |
Db        128 DLEDACNRRLGLVGTGRRLGLKGYAREVLGMAMEKPRRVTMSNWEKRELDPAQVEYACID 187

Qy        268 AYASWHL 274
              || |: |
Db        188 AYVSYKL 194


RESULT 8
Q9VE86
ID   Q9VE86      PRELIMINARY;      PRT;   346 AA.
AC   Q9VE86;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG7670 protein.
GN   CG7670.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
```

```
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003721; AAF55541.1; -.
DR   FlyBase; FBgn0038608; CG7670.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   SMART; SM00474; 35EXOc; 1.
SQ   SEQUENCE   346 AA;  39398 MW;  3150A9B4EFB74EB0 CRC64;

  Query Match             15.4%;  Score 229.5;  DB 5;  Length 346;
  Best Local Similarity   26.9%;  Pred. No. 5.5e-12;
  Matches   66;  Conservative   43;  Mismatches  123;  Indels   13;  Gaps    5;

Qy         41 TVQATTSVHGHEEDPNQIPNNIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILY 100
              |   |  :   ||:|   |     :| ||  |          : :    | ::: | | |
Db         68 TPDVTEKLAMEEENP---PKRRSSRLTRSTRSMAEDGSPSPEKEKPEKLPFIKYKGAIKY 124

Qy        101 SKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNY 160
              : ::   |  ::: :: ::||  :  :  |:||  ||: |  ||| | :|||||
Db        125 FTESQDIAASADDVLQWVEKQKDE--VVPMAFDMEWPFSFQTG--PGKSAVIQICVDEKC 180

Qy        161 CDVMHIFH-SGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDY-----GVSIKDVEDLSD 214
              | :  : :   :| :|  ||     :   |: |  |  ||  |:        |:   ||
Db        181 CYIYQLTNVKKLPAALVALINHPKVRLHGVNIKNDFRKLARDFPEVTAEPLIEKCVDLGL 240

Qy        215 LANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHL 274
                |:      :| |  ||  :  | : |  ::|:  |   || : || ||| | |    :
Db        241 WCNEVCETGGRWSLERLTNFIAKKAMDKSKKVRMSKWHVIPLDENQLMYAAIDVYIGQVI 300

Qy        275 YKVLK 279
              |: |:
Db        301 YRELE 305


RESULT 9
Q8S2J5
ID   Q8S2J5      PRELIMINARY;      PRT;   501 AA.
AC   Q8S2J5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   OSJNBb0021A09.10 protein.
GN   OSJNBB0021A09.10.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT   clone:OSJNBb0021A09.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003218; BAB89458.1; -.
SQ   SEQUENCE   501 AA;  56404 MW;  153AB72F4064FA99 CRC64;

  Query Match             15.3%;  Score 227.5;  DB 10;  Length 501;
  Best Local Similarity   35.6%;  Pred. No. 1.4e-11;
  Matches   58;  Conservative   28;  Mismatches   70;  Indels    7;  Gaps    5;

Qy        114 LIKVLDTKRDESGIAFVGLDIEWRP-SFRKGVLPGKVATVQICVDSNYCDVMHIFHSG-I 171
              | ::|   |       ||||:|||| ::  |  || || :|||     | :  | |:  :
Db         38 LDEILRIHRRRLNYLVVGLDVEWRPATYYHG--PGPVAVLQICV-GRRCLIFQILHADYV 94

Qy        172 PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGD--KKWGLA 229
              | ||   : |     ||:|| |  ||   .:|: :::| ||  || | ||    :  ||
Db         95 PDSLFDFLADGRFTFVGVGIHDDVAKLRSHHGLEVENVVDLRYLAAQTIGKPALRSAGLQ 154

Qy        230 SLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASW 272
              |   ::     ||  :|:  |: : |: :|: ||  ||:||:
Db        155 GLVREVMGVWAPKPYHVRVSAWDSWNLTPEQVMYACADAFASF 197


RESULT 10
Q961E1
ID   Q961E1      PRELIMINARY;      PRT;   353 AA.
AC   Q961E1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   GM14514p.
GN   CG7670.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY051647; AAK93071.1; -.
DR   FlyBase; FBgn0038608; CG7670.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
SQ   SEQUENCE   353 AA;  40248 MW;  F9836CB263868AB4 CRC64;

  Query Match             14.9%;  Score 221.5;  DB 5;  Length 353;
  Best Local Similarity   26.5%;  Pred. No. 2.9e-11;
  Matches   65;  Conservative   43;  Mismatches  124;  Indels   13;  Gaps    5;

Qy         41 TVQATTSVHGHEEDPNQIPNNIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILY 100
              |   |  :   ||:|   |     :| ||  |          : :    | ::: | | |
Db         75 TPDVTEKLAMEEENP---PKRRSSRLTRSTRSMAEDGSPSPEKEKPEKLPFIKYKGAIKY 131

Qy        101 SKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNY 160
              : ::   |  ::: :: :: |  :  :  |:||  ||: |  ||| | :|||||
Db        132 FTESQDIAASADDVLQWVEKQKAE--VVPMAFDMEWPFSFQTG--PGKSAVIQICVDEKC 187

Qy        161 CDVMHIFH-SGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDY-----GVSIKDVEDLSD 214
              | :  : :   :| :|  ||     :   |: |  |  ||  |:        |:   ||
Db        188 CYIYQLTNVKKLPAALVALINHPKVRLHGVNIKNDFRKLARDFPEVTAEPLIEKCVDLGL 247

Qy        215 LANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHL 274
                |:      :| |  ||  :  | : |  ::|:  |   || : || ||| | |    :
Db        248 WCNEVCETGGRWSLERLTNFIAKKAMDKSKKVRMSKWHVIPLDENQLMYAAIDVYIGQVI 307

Qy        275 YKVLK 279
              |: |:
Db        308 YRELE 312


RESULT 11
Q9VGN7
ID   Q9VGN7      PRELIMINARY;      PRT;   583 AA.
AC   Q9VGN7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CG6744 protein (AT25352p).
GN   CG6744.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
```

```
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003691; AAF54639.2; -.
DR   EMBL; AY089433; AAL90171.1; -.
DR   FlyBase; FBgn0037901; CG6744.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   SMART; SM00474; 35EXOc; 1.
SQ   SEQUENCE   583 AA;  66866 MW;  562984B177952AD0 CRC64;

  Query Match             14.3%;  Score 212.5;  DB 5;  Length 583;
  Best Local Similarity   33.5%;  Pred. No. 3.5e-10;
  Matches   61;  Conservative   23;  Mismatches   65;  Indels   33;  Gaps    6;

Qy        109 KRAMQLIKVLDTKRDESGIAFVGLDIEW-------RPSFRKGVLPGKVATVQICVDSNYC 161
                |   |  |||            | | ||       ||          || :|:      |
Db         71 KNHCQTFKVL------------GFDCEWITVGGSRRP----------VALLQLSSHRGLC 108

Qy        162 DVMHIFH-SGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKI 220
                :  : |   ||| |: |:|| :::|||:    |::|| ||||| :    ||  |   :
Db        109 ALFRLCHMKQIPQDLRELLEDDSVIKVGVAPQEDAMKLSHDYGVGVASTLDLRFLC--VM 166

Qy        221 GGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLY-KVLK 279
                | |  ||  |::| :   | |  |:   |||   |  :|| ||| ||  :  :| |: :
Db        167 AGHKPEGLGKLSKTHLNYTLDKHWRLACSNWEAKTLEPKQLDYAANDALMAVAIYQKLCR 226

Qy        280 DL 281
                ||
Db        227 DL 228


RESULT 12
Q9C7K6
ID   Q9C7K6      PRELIMINARY;      PRT;   582 AA.
AC   Q9C7K6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   3'-5' exonuclease, putative.
GN   F14G9.8.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzerg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
DR   EMBL; AC069159; AAG50917.1; -.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   InterPro; IPR002114; HPr_SerP_site.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   SMART; SM00474; 35EXOc; 1.
DR   PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
KW   Exonuclease.
SQ   SEQUENCE   582 AA;  66467 MW;  34D752473F786680 CRC64;

  Query Match             12.0%;  Score 179.5;  DB 10;  Length 582;
  Best Local Similarity   28.2%;  Pred. No. 2.8e-07;
  Matches   53;  Conservative   36;  Mismatches   66;  Indels   33;  Gaps    7;

Qy        126 GIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSN--YCDVMHIFHSG---IPQSLQHLIE 180
                |   ||:| ||:|:: ||    ||: :||  |:     |:: :::     :   | |:::
Db        377 GCRVVGIDCEWKPNYIKGSKQNKVSIMQIGSDTKIFILDLIKLYNDASEILDNCLSHILQ 436

Qy        181 DSTLVKV-------------GIGIDGDSVKLFHDYGVSIKDVE------DLSDLANQKI 220
                  : :|:             |     |  :|   ||  :|  |      |: :: |:
Db        437 SKSTLKLVSLTEDYPDHKLSSGYNFQCDIKQLALSYG-DLKCFERYDMLLDIQNVFNEPF 495

Qy        221 GGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKD 280
                |     ||| ||: ::   | |    |  :|| |||: ||:||| ||    |::: ::|
Db        496 G-----GLAGLTKKILGVSLNKTR--RNSDWEQRPLSQNQLEYAALDAAVLIHIFRHVRD 548

Qy        281 LPDAVSGS 288
                |   | |
Db        549 HPPHDSSS 556


RESULT 13
Q8SA18
ID   Q8SA18      PRELIMINARY;      PRT;   391 AA.
```

```
AC   Q8SA18;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   P0684E06.27 protein.
GN   P0684E06.27.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0684E06.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003291; BAB85281.1; -.
SQ   SEQUENCE   391 AA;  41510 MW;  9D5BD111CBE05C2D CRC64;

  Query Match           11.7%;  Score 174.5;  DB 10;  Length 391;
  Best Local Similarity   28.0%;  Pred. No. 4.5e-07;
  Matches   53;  Conservative   32;  Mismatches   89;  Indels   15;  Gaps    6;

Qy         95 GGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG-----KV 149
              |  |    | |     | | |  :  :| : |:  ||:| |||       ||      ||
Db        194 GDTIPIHTTITSSHSLAAQFINEIARERPQGGL-IVGIDTEWRTDH----LPNGKTCYKV 248

Qy        150 ATVQICVDSNYCDVMHIFHSG--IPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIK 207
              | :|:||    | :  |: :|  :|  |   : | ::  ||: :: |  :| :|  : :
Db        249 AVLQLCV-GRRCLLFQIYQAGNMVPHELAEFLADPSVREVGVAVNNDMQRLANDCNLRVA 307

Qy        208 DVEDLSDLANQKIGGDK--KWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAA 265
                ||   |  :|  :  : ||  |  |::   : |   |    |    |: :|: ||
Db        308 CAVDLRYAAAAVLGQPELARAGLKRLALTVMGAHMEKEKNITKSRWGEPTLTWEQVNYAC 367

Qy        266 TDAYASWHL 274
                ||| |: :
Db        368 IDAYVSYEI 376


RESULT 14
Q93VS2
ID   Q93VS2      PRELIMINARY;      PRT;   494 AA.
AC   Q93VS2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   P0487H02.21 protein (P0682B08.2 protein).
GN   P0487H02.21 OR P0682B08.2.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0487H02.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0682B08.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002883; BAB67861.1; -.
DR   EMBL; AP003578; BAB60932.1; -.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
SQ   SEQUENCE   494 AA;  51497 MW;  8F83F79405C2BD1E CRC64;

  Query Match           11.6%;  Score 173.5;  DB 10;  Length 494;
  Best Local Similarity   24.3%;  Pred. No. 7.5e-07;
  Matches   69;  Conservative   42;  Mismatches  126;  Indels   47;  Gaps   11;

Qy         13 EEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHG----------------HEEDPN 56
              :|||:      |:    |  |||       |  : |                |   |:
Db        204 DEELMG----SATGRARRPPSSSMGVGGEATETGGIGGGPPAPTSSSSPPPPPPHRPSPH 259

Qy         57 QIPNNIRRQLPRSITSSTSYKRPPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIK 116
                |:   | : ||          ||          : ||   :   | |     | : ::
Db        260 HTPHFFDRSIDRSA---------PLLLQDVDGGCTVSFGSAAI-DTTVTSDAAAADEWVR 309

Qy        117 -----VLDTKRDESGIAFVGLDIEWRP--SFRKGVLPGKVATVQICVDSNYCDVMHIFHS 169
                    | |   |:  |||| ||:|       | |  || :|:|   : | :: : |
Db        310 RVRASAATTPRGGGGL-LVGLDCEWKPCDHLWPAVAP-TVAILQLCAGDS-CLILQLLHV 366

Qy        170 G----IPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIG---G 222
                 :|  :  |: | ::  |||||  :: ||   |||       || |: ::::|   |
Db        367 AGARRVPPLVGDLLADPSVRLVGIGIGENAAKLADGYGVRCAAPVDLEDVCDRRLGRLPG 426

Qy        223 DKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
               :: ||      ::   : ||  :   :||   |   |::|| :
Db        427 ARRLGLKGYVREVLGLTMEKPMDVTRSDWERRHLDAAQVRYACS 470


RESULT 15
Q9C2I6
ID   Q9C2I6      PRELIMINARY;      PRT;   699 AA.
AC   Q9C2I6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Related to WERNER syndrome helicase.
GN   93G11.60.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL513443; CAC28658.1; -.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   SMART; SM00474; 35EXOc; 1.
SQ   SEQUENCE   699 AA;  77383 MW;  42FE96D430EAB1CA CRC64;

  Query Match           10.7%;  Score 159.5;  DB 3;  Length 699;
  Best Local Similarity   26.0%;  Pred. No. 2.1e-05;
  Matches   52;  Conservative   42;  Mismatches   73;  Indels   33;  Gaps    9;

Qy         97 RILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGK-VATVQIC 155
              :: | ::|| ::   |    :| |        :|||:|| | ::   | : |: :||
Db        162 KVHYCRSATTTERVLKQYF--MDEK-------ILGLDLEWEISAKESHGPRQNVSVIQIA 212

Qy        156 VDSNYCDVMHIFHSGI-PQS-------LQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIK 207
              :    : |||  : |:        |: :|||: :|| |: | ||  :|    |:  |
Db        213 SEKR----IGIFHISLYPRKDELASPLLKQIIEDADVVKAGVWIMGDCTRLKTFLGIEAK 268

Qy        208 DVEDLSDL---------ANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSK 258
               : :|| |           |:   :   ||:| : ::   : |   :|   |   ||:
Db        269 GIYELSHLYKLVKYSASGEHKLVNRRLVPLATLVKEVLQLPMFK-GAVRTSEWS-KPLNM 326

Qy        259 QQLQYAATDAYASWHLYKVL 278
```

```
                    |: |: :|||| |: ::
Db        327 DQILYSGSDAYAGVQLFAMM 346
```

Search completed: June 6, 2003, 10:24:14
Job time : 35 secs

OM protein - protein search, using sw model

Run on: June 6, 2003, 10:22:40 ; Search time 15 Seconds
(without alignments)
564.920 Million cell updates/sec

Title: US-09-896-186B-24
Perfect score: 1491
Sequence: 1 MSSSNWIDDAFTEEELLAID.....YASWHLYKVLKDLPDAVSGS 288

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	352	23.6	1432	3	US-08-781-891-71	Sequence 71, Appl
2	339	22.7	1401	4	US-09-127-670-6	Sequence 6, Appli
3	330	22.1	1401	3	US-08-781-891-206	Sequence 206, App
4	93	6.2	3079	5	PCT-US94-00198-4	Sequence 4, Appli
5	89	6.0	930	4	US-09-134-001C-5314	Sequence 5314, Ap
6	87.5	5.9	1375	4	US-09-722-139-2	Sequence 2, Appli
7	87.5	5.9	1375	4	US-09-721-832-2	Sequence 2, Appli
8	87.5	5.9	1375	4	US-09-721-689-2	Sequence 2, Appli
9	86.5	5.8	348	1	US-07-989-845-2	Sequence 2, Appli
10	86.5	5.8	348	5	PCT-US93-11298-2	Sequence 2, Appli
11	86.5	5.8	480	4	US-08-987-367-2	Sequence 2, Appli
12	82.5	5.5	480	4	US-08-987-367-4	Sequence 4, Appli
13	82	5.5	333	4	US-09-134-001C-4528	Sequence 4528, Ap
14	80	5.4	574	2	US-08-907-166-12	Sequence 12, Appl
15	79.5	5.3	431	4	US-08-845-258-34	Sequence 34, Appl
16	79.5	5.3	431	4	US-08-990-571-34	Sequence 34, Appl
17	79.5	5.3	431	4	US-08-723-142A-34	Sequence 34, Appl
18	79.5	5.3	431	4	US-09-528-784A-34	Sequence 34, Appl
19	79.5	5.3	616	4	US-09-388-743-14	Sequence 14, Appl
20	79	5.3	927	3	US-08-895-601-6	Sequence 6, Appli
21	78.5	5.3	260	4	US-09-134-001C-5363	Sequence 5363, Ap
22	78.5	5.3	637	4	US-09-134-001C-5658	Sequence 5658, Ap
23	78.5	5.3	817	1	US-07-640-029-2	Sequence 2, Appli
24	78.5	5.3	822	1	US-07-997-133-1	Sequence 1, Appli
25	78.5	5.3	822	1	US-07-921-807B-4	Sequence 4, Appli
26	78.5	5.3	822	1	US-08-459-296-2	Sequence 2, Appli
27	78.5	5.3	822	1	US-08-441-944A-4	Sequence 4, Appli
28	78.5	5.3	822	2	US-08-451-822A-12	Sequence 12, Appl
29	78.5	5.3	822	4	US-08-439-992A-2	Sequence 2, Appli
30	78.5	5.3	822	4	US-08-323-430-12	Sequence 12, Appl
31	77.5	5.2	247	4	US-09-216-295-8	Sequence 8, Appli
32	76	5.1	1196	4	US-08-881-706-2	Sequence 2, Appli
33	75.5	5.1	407	4	US-09-271-438A-11	Sequence 11, Appl
34	75.5	5.1	701	4	US-09-132-028-2	Sequence 2, Appli
35	75.5	5.1	1056	2	US-08-687-289A-7	Sequence 7, Appli
36	75.5	5.1	1056	2	US-08-687-289A-8	Sequence 8, Appli
37	75.5	5.1	1058	2	US-08-687-289A-5	Sequence 5, Appli
38	75.5	5.1	1078	1	US-08-485-588-7	Sequence 7, Appli
39	75.5	5.1	1078	1	US-08-484-565-7	Sequence 7, Appli
40	75.5	5.1	1078	2	US-08-480-751-7	Sequence 7, Appli
41	75.5	5.1	1078	2	US-08-943-986-7	Sequence 7, Appli
42	75.5	5.1	1078	3	US-08-353-784-7	Sequence 7, Appli
43	75.5	5.1	1078	3	US-08-484-719B-7	Sequence 7, Appli
44	75.5	5.1	1078	4	US-08-484-159-7	Sequence 7, Appli
45	75.5	5.1	1088	1	US-08-485-588-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-781-891-71
; Sequence 71, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1432 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-781-891-71

Query Match 23.6%; Score 352; DB 3; Length 1432;
Best Local Similarity 37.7%; Pred. No. 8.3e-30;
Matches 75; Conservative 41; Mismatches 75; Indels 8; Gaps 3;

Qy 88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147

```
Db          43 DLPFLEFTGSIVYSYDASDCSFLSE------DISMSLSDGDVVGFDMEWPPLYNRGKL-G 95

Qy         148 KVATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSI 206
               ||| :|:||  : | : |:    : || |: |:|:  : | |:||:||  ||  |: : :
Db          96 KVALIQLCVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKL 155

Qy         207 KDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
               |:  :|:|:||:|:   : | | || : |: |:|||   ||  ||  :||:: |  ||||
Db         156 KNFVELTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLYAAT 215

Qy         267 DAYASWHLYKVLKDLPDAV 285
               |||| : :|: |: | | |
Db         216 DAYAGFIIYRNLEILDDTV 234


RESULT 2
US-09-127-670-6
; Sequence 6, Application US/09127670
; Patent No. 6228583
; GENERAL INFORMATION:
;  APPLICANT: Massachusetts Institute for Technology
;  APPLICANT: Leonard P. Guarente
;  APPLICANT: David A. Sinclair
;  APPLICANT: David B. Lombard
;  TITLE OF INVENTION: ASSAYS FOR COMPOUNDS WHICH EXTEND LIFE
;  TITLE OF INVENTION: SPAN
;  FILE REFERENCE: MIT-7720pA
;  CURRENT APPLICATION NUMBER: US/09/127,670
;  CURRENT FILING DATE: 1998-07-31
;  EARLIER APPLICATION NUMBER: US 60/054,629
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 1401
;   TYPE: PRT
;   ORGANISM: Murine
US-09-127-670-6

  Query Match             22.7%;  Score 339;  DB 4;  Length 1401;
  Best Local Similarity   36.6%;  Pred. No. 2.2e-28;
  Matches   74;  Conservative   42;  Mismatches   72;  Indels   14;  Gaps    5;

Qy          88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
               | | : | | |:||  |::       | : :  :  :  :  || |:|| | ::    ||
Db          37 NLPFLEFPGSIVYSYEASD----CSFLSEDISMRLSDGDV--VGFDMEWPPIYK----PG 86

Qy         148 K---VATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG 203
               |   || :|:||  | | : ||    : || |: |:|: :: | |:||:||  ||  |:
Db          87 KRSRVAVIQLCVSENKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRDFD 146

Qy         204 VSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQY 263
               | ::   :|:|:||:|:   : | |  | : :: |:|||   ||  ||  :||:: |  |
Db         147 VKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLY 206

Qy         264 AATDAYASWHLYKVLKDLPDAV 285
               |||||||   :|: | :| | |
Db         207 AATDAYAGLIIYQKLGNLGDTV 228


RESULT 3
US-08-781-891-206
; Sequence 206, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  206:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1401 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-781-891-206

  Query Match             22.1%;  Score 330;  DB 3;  Length 1401;
  Best Local Similarity   36.0%;  Pred. No. 2.3e-27;
  Matches   72;  Conservative   43;  Mismatches   71;  Indels   14;  Gaps    5;

Qy          88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
               | | : | | |:||  |::       | : :  :  :  :  || |:|| | ::    ||
Db          37 NLPFLEFPGSIVYSYEASD----CSFLSEDISMRLSDGDV--VGFDMEWPPIYK----PG 86

Qy         148 K---VATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG 203
               |   || :|:||  : | : ||    : || |: |:|: :: | |:||:||  ||  |:
Db          87 KRSRVAVIQLCVSESKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRDFD 146

Qy         204 VSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQY 263
               | ::   :|:|:||:|:   : | |  | : :: |:|||   ||  ||  :||:: |  |
Db         147 VKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLY 206

Qy         264 AATDAYASWHLYKVLKDLPD 283
               |||||||   :|: | :| |
Db         207 AATDAYAGLIIYQKLGNLGD 226


RESULT 4
PCT-US94-00198-4
; Sequence 4, Application PC/TUS9400198
;  GENERAL INFORMATION:
;    APPLICANT:  Schering Corp.
;    TITLE OF INVENTION:  RAS Associated GAP Proteins
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schering Corp.
;      STREET:  1 Girald Farms
;      CITY:  Madison
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  94304-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  6.0.8
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/00198
```

```
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/004,824
;       FILING DATE:  15-JAN-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lunn, Paul G.
;       REGISTRATION NUMBER:  32,743
;       REFERENCE/DOCKET NUMBER:  DX0352 PCT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (201)822-7255
;       TELEFAX:  (201)822-7039
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3079 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     ORIGINAL SOURCE:
;       ORGANISM:  Saccharomyces cerevisiae
PCT-US94-00198-4

  Query Match             6.2%;  Score 93;  DB 5;  Length 3079;
  Best Local Similarity   19.3%;  Pred. No. 2;
  Matches   57;  Conservative   43;  Mismatches  112;  Indels   84;  Gaps   10;

Qy          9 DAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPR 68
              | | :   | || :       :   |:       :    :  | :|  ::   : |   |
Db       1976 DDFYKTTFLLIDDVLGQLGQPKMEFSN-------EIPIYIREHMDDYPELYEFMNRHAFR 2028

Qy         69 SITSSTSY----------KRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIK-- 116
              :| :||:|          :  |:      ||           |    ::|  | : ::
Db       2029 NIETSTAYSPSVHESTSSEGIPIITLTMSNF----------SDRHVDIDTVAYKFLQIY 2077

Qy        117 ----------VLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHI 166
                       ::|    : |    |||:    |   |:|| :||  : |:   | :|
Db       2078 ARIWTTKHCLIIDCTEFDEG----GLDMRKFISLVMGLLP-EVAP-KNCIGCYYFNVNET 2131

Qy        167 F---------------HSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 211
              |               | ||    :   |  |:|  :|| |  :|:  |  ||: |:
Db       2132 FMDNYGKCLDKDNVYVSSKIPHYFINSNSDEGLMK-SVGITGQGLKVLQDIRVSLHDITL 2190

Qy        212 LSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATD 267
                :  |:                    |  :::|:  |  | :   ||   |
Db       2191 YDEKRNR---------------------FTPVSLKIGDIYFQVLHETPRQYKIRD 2224


RESULT 5
US-09-134-001C-5314
; Sequence 5314, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5314
;   LENGTH: 930
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5314

  Query Match             6.0%;  Score 89;  DB 4;  Length 930;
  Best Local Similarity   23.6%;  Pred. No. 0.84;
  Matches   71;  Conservative   45;  Mismatches  117;  Indels   68;  Gaps   16;

Qy          1 MSSSNWIDDAFTEEELLAID---AIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQ 57
              : |||   |   :     |:   :|: | |   |  | : :  ::: |     ||:  :
Db        150 VESSNSSMDTAQQPSHTTINSEASIQTSDNEENSRVSDFANSKIIESNTE-SNKEENTIE 208

Qy         58 IPNNIRRQLPRSITSS-TSYKRF----------------------PLSRCRARNFPAMR 93
              || :|    ||||  :|||                          |||   |:  |: :
Db        209 QPNKVRED---SITSQPSSYKNIDEKISNQDELLNLPINEYENKVRPLSTTSAQ--PSSK 263

Qy         94 FGGRILYSKTATEVDKRAMQLIKVLDTK-----RDESGI--AFVGLDIEWRPSFR--KGV 144
                |:  :: | |       |||| |         |  ||  |   :: :  :|     |
Db        264 ---RVTVNQLAAEQGSNVNHLIKVTDQSITEGYDDSDGIIKAHDAENLIYDVTFEVDDKV 320

Qy        145 LPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVK----LFH 200
              |   || |  ::   |:   |   ||:     |:|::   :  |   :: |     |
Db        321 KSGDTMTVNIDKNTVPSDLTDSF--AIPK-----IKDNSGEIIATGTYDNTNKQITYTFT 373

Qy        201 DYGVSIKDVE---------DLSDLANQ--KIGGDKKWGLASLTETLVCKELLKPNRIRLG 249
              ||    ::::         | | : |   |:  : |  |:|: :|:   |  |||  |
Db        374 DYVDKYENIKAHLKLTSYIDKSKVPNNNTKLDVEYKTALSSVNKTITV-EYQKPNENRTA 432

Qy        250 N 250
              |
Db        433 N 433


RESULT 6
US-09-722-139-2
; Sequence 2, Application US/09722139
; Patent No. 6355471
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6355471el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1055
;  CURRENT APPLICATION NUMBER: US/09/722,139
;  CURRENT FILING DATE:  2000-11-24
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1375
;   TYPE: PRT
;   ORGANISM: Human
US-09-722-139-2

  Query Match             5.9%;  Score 87.5;  DB 4;  Length 1375;
  Best Local Similarity   19.0%;  Pred. No. 2.3;
  Matches   58;  Conservative   52;  Mismatches  116;  Indels   79;  Gaps   13;

Qy         13 EEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPRSITS 72
              :|| ||         :  :::       | : | ::   ||        :|::    : |
Db        935 KEEQLA--------QYQANANQLQKLQATFEFTANIARQEE-------KVRKKEKEILES 979

Qy         73 STSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGL 132
                 :|  | |  ||      |       :|   ||::::  :|  :    |::||:    |
Db        980 REKQQREALERALAR--LERRHSALQRHSTLGTEIEEQRQKLASLNSGSREQSGLQ-ASL 1036

Qy        133 DIEWRPSFRK-------------------------GVLPGKVATVQICVDSNYCDVMH 165
              : | : :  |                        | | ||||:  : | :    ::
Db       1037 EAE-QEALEKDQERLEYEIQQLKQKIYEVDGVQKDHHGTLEGKVASSSLPVSAEKSHLVP 1095

Qy        166 I----FHSGIPQSLQHLIED-STLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKI 220
              :     :: | : :|  ::|   ::  |      |::|             |   | |  |
Db       1096 LMDARINAYIEEEVQRRLQDLHRVISEGCSTSADTMK------------DNEKLHNGTI 1142

Qy        221 GGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKD 280
                |:    |   |:|  : :|:      |   :|| :| |   : |   |        :
Db       1143 QRKLKY---ELCRDLLCVLMPEPDAAACAN---HPLLQQDLVQLSLD----WK-----TE 1187

Qy        281 LPDAV 285
              :|| |
Db       1188 IPDLV 1192
```

```
RESULT 7
US-09-721-832-2
; Sequence 2, Application US/09721832
; Patent No. 6399346
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6399346el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1055
;  CURRENT APPLICATION NUMBER: US/09/721,832
;  CURRENT FILING DATE:  2000-11-24
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1375
;   TYPE: PRT
;   ORGANISM: Human
US-09-721-832-2

  Query Match             5.9%;  Score 87.5;  DB 4;  Length 1375;
  Best Local Similarity   19.0%;  Pred. No. 2.3;
  Matches   58;  Conservative   52;  Mismatches  116;  Indels   79;  Gaps   13;

Qy         13 EEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPRSITS 72
              :|| ||         :  :::       | : | ::   ||        :|::    : |
Db        935 KEEQLA--------QYQANANQLQKLQATFEFTANIARQEE-------KVRKKEKEILES 979

Qy         73 STSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGL 132
                 :|  | |  ||     |       :|   ||:::: :|  :    |::||:    |
Db        980 REKQQREALERALAR--LERRHSALQRHSTLGTEIEEQRQKLASLNSGSREQSGLQ-ASL 1036

Qy        133 DIEWRPSFRK--------------------------GVLPGKVATVQICVDSNYCDVMH 165
              : | : :  |                          | | ||||:  : | :    ::
Db       1037 EAE-QEALEKDQERLEYEIQQLKQKIYEVDGVQKDHHGTLEGKVASSSLPVSAEKSHLVP 1095

Qy        166 I----FHSGIPQSLQHLIED-STLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKI 220
              :     :: | : :|  ::|   ::  |     |::|             |   | | |
Db       1096 LMDARINAYIEEEVQRRLQDLHRVISEGCSTSADTMK------------DNEKLHNGTI 1142

Qy        221 GGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKD 280
                  |:    |   |:|  : :|:     |  :|| :| |   : |    |       :
Db       1143 QRKLKY---ELCRDLLCVLMPEPDAAACAN---HPLLQQDLVQLSLD----WK-----TE 1187

Qy        281 LPDAV 285
              :|| |
Db       1188 IPDLV 1192


RESULT 8
US-09-721-689-2
; Sequence 2, Application US/09721689
; Patent No. 6440685
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6440685el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1055
;  CURRENT APPLICATION NUMBER: US/09/721,689
;  CURRENT FILING DATE:  2000-11-24
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1375
;   TYPE: PRT
;   ORGANISM: Human
US-09-721-689-2

  Query Match             5.9%;  Score 87.5;  DB 4;  Length 1375;
  Best Local Similarity   19.0%;  Pred. No. 2.3;
  Matches   58;  Conservative   52;  Mismatches  116;  Indels   79;  Gaps   13;

Qy         13 EEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPRSITS 72
              :|| ||         :  :::       | : | ::   ||        :|::    : |
Db        935 KEEQLA--------QYQANANQLQKLQATFEFTANIARQEE-------KVRKKEKEILES 979

Qy         73 STSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGL 132
                 :|  | |  ||     |       :|   ||:::: :|  :    |::||:    |
Db        980 REKQQREALERALAR--LERRHSALQRHSTLGTEIEEQRQKLASLNSGSREQSGLQ-ASL 1036

Qy        133 DIEWRPSFRK--------------------------GVLPGKVATVQICVDSNYCDVMH 165
              : | : :  |                          | | ||||:  : | :    ::
Db       1037 EAE-QEALEKDQERLEYEIQQLKQKIYEVDGVQKDHHGTLEGKVASSSLPVSAEKSHLVP 1095

Qy        166 I----FHSGIPQSLQHLIED-STLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKI 220
              :     :: | : :|  ::|   ::  |     |::|             |   | | |
Db       1096 LMDARINAYIEEEVQRRLQDLHRVISEGCSTSADTMK------------DNEKLHNGTI 1142

Qy        221 GGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKD 280
                  |:    |   |:|  : :|:     |  :|| :| |   : |    |       :
Db       1143 QRKLKY---ELCRDLLCVLMPEPDAAACAN---HPLLQQDLVQLSLD----WK-----TE 1187

Qy        281 LPDAV 285
              :|| |
Db       1188 IPDLV 1192


RESULT 9
US-07-989-845-2
; Sequence 2, Application US/07989845
; Patent No. 5304472
; GENERAL INFORMATION:
;   APPLICANT:  Bass, Steven
;   APPLICANT:  Swartz, James
;   TITLE OF INVENTION:  METHOD OF CONTROLLING POLYPEPTIDE
;   TITLE OF INVENTION:  PRODUCTION IN BACTERIAL CELLS
;   NUMBER OF SEQUENCES:  31
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Genentech, Inc.
;     STREET:  460 Point San Bruno Blvd
;     CITY:  South San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94080-4990
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  patin (Genentech)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/989,845
;     FILING DATE:  19921120
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Hasak, Janet E.
;     REGISTRATION NUMBER:  28,616
;     REFERENCE/DOCKET NUMBER:  752
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  415/225-1896
;     TELEFAX:  415/952-9881
;     TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  348 amino acids
;     TYPE:  AMINO ACID
;     TOPOLOGY:  linear
US-07-989-845-2

  Query Match             5.8%;  Score 86.5;  DB 1;  Length 348;
  Best Local Similarity   20.9%;  Pred. No. 0.34;
```

```
  Matches   54;  Conservative   34;  Mismatches   73;  Indels   97;  Gaps   12;

Qy         86 ARNFPAMRFGGRILYSKTATEV--------DKRAM----------QLIKVLDTKRDESG 126
              | | | :: |  :|  ||  ::       |  |:          | | |:  : | ||
Db        107 AVNIPGLKSGELVLDGKTLGDIYLGKIKKWDDEAIAKLNPGLKLPSQNIAVV-RRADGSG 165

Qy        127 IAFV------GLDIEWRPSFRKGV---------------------LPGKVATVQICVDS 158
              :||       :: ||: :   |                     ||| :  |:
Db        166 TSFVFTSYLAKVNEEWKNNVGTGSTVKWPIGLGGKGNDGIAAFVQRLPGAIGYVEYA--- 222

Qy        159 NYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLS----- 213
               |    :: :: :                  |  ||  |    :   :     | |
Db        223 -YAKQNNLAYTKL----------------ISADGKPVSPTEENFANAAKGADWSKTFAQ 264

Qy        214 DLANQKIGGDKKWGLASLTETLVCKELLKPNR----IRLGNWEFYPLSKQ--QLQYAATD 267
              || |||  |:  | : | |  |: |:  || :    ::  :| :   :||   | ||:
Db        265 DLTNQK--GEDAWPITSTTFILIHKDQKKPEQGTEVLKFFDWAYKTGAKQANDLDYAS-- 320

Qy        268 AYASWHLYKVLKDLPDAV 285
                         |||:|
Db        321 ------------LPDSV 325


RESULT 10
PCT-US93-11298-2
; Sequence 2, Application PC/TUS9311298
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc.
;    TITLE OF INVENTION:  METHOD OF CONTROLLING POLYPEPTIDE PRODUCTION IN
;    TITLE OF INVENTION:  BACTERIA
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/11298
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  752
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  348 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
PCT-US93-11298-2

  Query Match             5.8%;  Score 86.5;  DB 5;  Length 348;
  Best Local Similarity   20.9%;  Pred. No. 0.34;
  Matches   54;  Conservative   34;  Mismatches   73;  Indels   97;  Gaps   12;

Qy         86 ARNFPAMRFGGRILYSKTATEV--------DKRAM----------QLIKVLDTKRDESG 126
              | | | :: |  :|  ||  ::       |  |:          | | |:  : | ||
Db        107 AVNIPGLKSGELVLDGKTLGDIYLGKIKKWDDEAIAKLNPGLKLPSQNIAVV-RRADGSG 165

Qy        127 IAFV------GLDIEWRPSFRKGV---------------------LPGKVATVQICVDS 158
              :||       :: ||: :   |                     ||| :  |:
Db        166 TSFVFTSYLAKVNEEWKNNVGTGSTVKWPIGLGGKGNDGIAAFVQRLPGAIGYVEYA--- 222

Qy        159 NYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLS----- 213
               |    :: :: :                  |  ||  |    :   :     | |
Db        223 -YAKQNNLAYTKL----------------ISADGKPVSPTEENFANAAKGADWSKTFAQ 264

Qy        214 DLANQKIGGDKKWGLASLTETLVCKELLKPNR----IRLGNWEFYPLSKQ--QLQYAATD 267
              || |||  |:  | : | |  |: |:  || :    ::  :| :   :||   | ||:
Db        265 DLTNQK--GEDAWPITSTTFILIHKDQKKPEQGTEVLKFFDWAYKTGAKQANDLDYAS-- 320

Qy        268 AYASWHLYKVLKDLPDAV 285
                         |||:|
Db        321 ------------LPDSV 325


RESULT 11
US-08-987-367-2
; Sequence 2, Application US/08987367
; Patent No. 6399859
;  GENERAL INFORMATION:
;    APPLICANT:  Nichols, Scott
;    APPLICANT:  Dhugga, Kanwarpal S.
;    APPLICANT:  Singletary, George William
;    APPLICANT:  Saunders, Court
;    APPLICANT:  Witcher, Derrick
;    APPLICANT:  Bruce, Wesley B.
;    APPLICANT:  Sims, Lynne
;    APPLICANT:  Lu, Guihua
;    APPLICANT:  Zhong, Gan-Yuan
;    TITLE OF INVENTION:  PLANT URIDINE DIPHOSPHATE-GLUCOSE
;    TITLE OF INVENTION:  DEHYDROGENASE GENES, PROTEINS, AND USES THEREOF
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/987,367
;      FILING DATE:  10-DEC-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jones, Phil B.C.
;      REGISTRATION NUMBER:  38,195
;      REFERENCE/DOCKET NUMBER:  750027.404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  480 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-987-367-2

  Query Match             5.8%;  Score 86.5;  DB 4;  Length 480;
  Best Local Similarity   22.1%;  Pred. No. 0.56;
  Matches   43;  Conservative   32;  Mismatches   63;  Indels   57;  Gaps    9;

Qy         81 LSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSF 140
              : :|| ||         : :|   |:|:|              |: | || ::    |:
Db         58 VKQCRGRN---------LFFS---TDVEKHVA-----------EADIIFVSVN---TPTK 91
```

```
Qy        141 RKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFH 200
              :|:  || |     |  | :      : : :| : ::| ||
Db         92 TRGLGAGKAA------DLTYWESAARMIADVSKSDKIVVEKST---------------- 128

Qy        201 DYGVSIKDVEDLSDLANQKIGGDKKWGLAS---LTETLVCKELLKPNRIRLGNWEFYPLS 257
                | :|  | :  :      |     |::   | |    ::| ||:|: :|  |  |
Db        129 ---VPVKTAEAIEKILTHNSKGINYQILSNPEFLAEGTAIEDLFKPDRVLIGGRE-TPEG 184

Qy        258 KQQLQYAATDAYASW 272
              :: :| |   | || |
Db        185 RKAVQ-ALKDVYAHW 198


RESULT 12
US-08-987-367-4
; Sequence 4, Application US/08987367
; Patent No. 6399859
;  GENERAL INFORMATION:
;    APPLICANT:  Nichols, Scott
;    APPLICANT:  Dhugga, Kanwarpal S.
;    APPLICANT:  Singletary, George William
;    APPLICANT:  Saunders, Court
;    APPLICANT:  Witcher, Derrick
;    APPLICANT:  Bruce, Wesley B.
;    APPLICANT:  Sims, Lynne
;    APPLICANT:  Lu, Guihua
;    APPLICANT:  Zhong, Gan-Yuan
;    TITLE OF INVENTION:  PLANT URIDINE DIPHOSPHATE-GLUCOSE
;    TITLE OF INVENTION:  DEHYDROGENASE GENES, PROTEINS, AND USES THEREOF
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/987,367
;      FILING DATE:  10-DEC-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jones, Phil B.C.
;      REGISTRATION NUMBER:  38,195
;      REFERENCE/DOCKET NUMBER:  750027.404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  480 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-987-367-4

  Query Match             5.5%;  Score 82.5;  DB 4;  Length 480;
  Best Local Similarity   21.5%;  Pred. No. 1.6;
  Matches   42;  Conservative   32;  Mismatches   64;  Indels   57;  Gaps    9;

Qy         81 LSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSF 140
              : :|| ||         : :|   |:|:|              |: | || ::    |:
Db         58 VKQCRGRN---------LFFS---TDVEKHVA-----------EADIIFVSVN---TPTK 91

Qy        141 RKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFH 200
              :|:  || |     |  | :      : : :| : ::| ||
Db         92 TRGLGAGKAA------DLTYWESAARMIADVSKSDKIVVEKST---------------- 128

Qy        201 DYGVSIKDVEDLSDLANQKIGGDKKWGLAS---LTETLVCKELLKPNRIRLGNWEFYPLS 257
                | :|  | :  :      |     |::   | |    ::| ||:|: :|  |  |
Db        129 ---VPVKTAEAIEKILTHNSKGINYQILSNPEFLAEGTAIEDLFKPDRVLIGGRE-TPEG 184

Qy        258 KQQLQYAATDAYASW 272
              :: :| |   | || |
Db        185 RKAVQ-VLKDVYAHW 198


RESULT 13
US-09-134-001C-4528
; Sequence 4528, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4528
;   LENGTH: 333
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4528

  Query Match             5.5%;  Score 82;  DB 4;  Length 333;
  Best Local Similarity   21.1%;  Pred. No. 1;
  Matches   38;  Conservative   38;  Mismatches   68;  Indels   36;  Gaps    9;

Qy         99 LYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVL--PGKVATVQICV 156
              :| :  |::|:    |: : |:   :  :||  :       :|  |  || :  :   :
Db        163 IYKRLLTKIDE-PENLVIIGDSAGGQIALAFAQM-------LKKEQLSQPGHIVLISPVL 214

Qy        157 DSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSI-KDVEDLSDL 215
              |: : :     |:: ::  ||  |     ||||        | | :      | |
Db        215 DATFKN---------PEARKYEKEDPML-----GIDGSK------YLVELWAGDAPLDDY 254

Qy        216 ANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLY 275
                : || : ||  :| |:  || | |: ::      : |: : :::     |  :|:|
Db        255 KMSPMNGDLE-GLGHITLTVGTKETLYPDAVKFS----HMLNDKGIKHQFIPGYNLFHIY 309


RESULT 14
US-08-907-166-12
; Sequence 12, Application US/08907166
; Patent No. 5948666
; GENERAL INFORMATION:
;  APPLICANT: Callen, Walter
;  APPLICANT: Mather, Eric
;  TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
;  FILE REFERENCE: 09010/027001
;  CURRENT APPLICATION NUMBER: US/08/907,166
;  CURRENT FILING DATE: 1997-08-06
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12
;   LENGTH: 574
;   TYPE: PRT
;   ORGANISM: Aquifex pyrophilus
US-08-907-166-12

  Query Match             5.4%;  Score 80;  DB 2;  Length 574;
  Best Local Similarity   21.9%;  Pred. No. 3.9;
  Matches   33;  Conservative   30;  Mismatches   48;  Indels   40;  Gaps    6;

Qy        170 GIPQSLQHLIEDSTLVKVGIGIDGDSVKLF------HDYGV---SIKDVEDLSDLANQK- 219
              |: :::: | |:|  : :     || ::|       : | :    |:|:| |  | |::
```

```
Db         11 GLKKAIKRL-ENSPYLYLDTETTGDRIRLVQIGDEENTYVIDLYEIQDIEPLRKLINERG 69

Qy        220 -IGGDKKWGL----------ASLTETLVCKELLKPNRIRL-----------------GN 250
Db         70 IVGHNLKFDLKYLYRYGIFPSATFDTMIASYLLGYERHSLNHIVSNLLGYSMDKSYQTSD 129

Qy        251 WEFYPLSKQQLQYAATDAYASWHLYKVLKDL 281
Db        130 WGASVLSDAQLKYAANDVIVLRELFPKMRDM 160


RESULT 15
US-08-845-258-34
; Sequence 34, Application US/08845258
; Patent No. 6183976
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Sleath, Paul R.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;    TITLE OF INVENTION:  AND TREATMENT OF B. MICROTI INFECTION
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED AND BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/845,258
;      FILING DATE:  24-APR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.426C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  431 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-845-258-34

  Query Match             5.3%;  Score 79.5;  DB 4;  Length 431;
  Best Local Similarity   22.4%;  Pred. No. 2.9;
  Matches   66;  Conservative   35;  Mismatches  106;  Indels   87;  Gaps   17;

Qy          9 DAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPR 68
Db         33 DNFTLDNPSAYEILRVSYNSNEFQVQS------------------PQNINNEMESSTPE 73

Qy         69 S-----ITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRD 123
Db         74 SNIIWVVHSDVIMKRF---NCKNR---------KSLSTHSLTEND--------ILKFGRI 113

Qy        124 ESGI--AFVGLDIEWRPSFRKG---VLPGKVATVQICVDSNYCDVMH-IFHSGIPQSLQH 177
Db        114 ELSVKCIIMGAGITASDLNLKGLGFISPDKQST-NVC---NYFEDMHESYHILDTQRASD 169

Qy        178 LIEDS---------TLVKVG--IGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKW 226
Db        170 CVSDDGADIDISNFDMVQDGNINSVDADSETCMANSGVTVNNTENVSNSEN--------F 221

Qy        227 G-LASL--TETLVCKELLKPNRIRLGNWEFYPL-----SKQQLQYAATDAYASW 272
Db        222 GKLKSLVSTTTPLCRICL------CGESDPGPLVTPCNCKGSLNYVHLECLRTW 269


Search completed: June  6, 2003, 10:24:59
Job time : 17 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 6, 2003, 10:24:21 ; Search time 21 Seconds
(without alignments)
1415.867 Million cell updates/sec

Title: US-09-896-186B-24
Perfect score: 1491
Sequence: 1 MSSSNWIDDAFTEEELLAID.....YASWHLYKVLKDLPDAVSGS 288

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	151	10.1	147	10	US-09-729-674-279	Sequence 279, App
2	93	6.2	2224	9	US-10-115-563-14	Sequence 14, Appl
3	89.5	6.0	930	10	US-09-815-242-11133	Sequence 11133, A
4	87.5	5.9	1885	10	US-09-920-346-2	Sequence 2, Appli
5	87	5.8	910	9	US-10-174-363-22	Sequence 22, Appl
6	86.5	5.8	346	10	US-09-741-669-380	Sequence 380, App
7	86.5	5.8	367	9	US-10-268-441-15	Sequence 15, Appl
8	86.5	5.8	373	9	US-10-268-441-14	Sequence 14, Appl
9	86.5	5.8	593	10	US-09-815-242-12306	Sequence 12306, A
10	86	5.8	1042	10	US-09-888-615-74	Sequence 74, Appl
11	85.5	5.7	1394	9	US-10-108-605-213	Sequence 213, App
12	83.5	5.6	319	9	US-09-738-626-6900	Sequence 6900, Ap
13	82.5	5.5	305	9	US-09-738-626-6005	Sequence 6005, Ap
14	81	5.4	389	9	US-10-174-363-50	Sequence 50, Appl
15	80	5.4	574	10	US-09-391-340-12	Sequence 12, Appl
16	80	5.4	574	10	US-09-948-369-12	Sequence 12, Appl
17	79.5	5.3	431	9	US-09-286-488-34	Sequence 34, Appl
18	79.5	5.3	431	10	US-09-737-178-34	Sequence 34, Appl
19	79	5.3	374	9	US-10-268-441-6	Sequence 6, Appli
20	79	5.3	733	9	US-10-097-534-16	Sequence 16, Appl
21	79	5.3	927	9	US-10-097-534-15	Sequence 15, Appl
22	78.5	5.3	764	10	US-09-925-302-714	Sequence 714, App
23	78.5	5.3	7968	9	US-10-077-130-5	Sequence 5, Appli
24	77	5.2	515	10	US-09-801-368-384	Sequence 384, App
25	76.5	5.1	822	9	US-09-757-415A-2	Sequence 2, Appli
26	76	5.1	393	10	US-09-815-242-11856	Sequence 11856, A
27	76	5.1	396	10	US-09-864-761-43232	Sequence 43232, A
28	76	5.1	1196	9	US-09-823-394-2	Sequence 2, Appli
29	75.5	5.1	1078	9	US-10-125-792-28	Sequence 28, Appl
30	75.5	5.1	1078	9	US-10-125-778-28	Sequence 28, Appl
31	75.5	5.1	1078	10	US-09-727-205-2	Sequence 2, Appli
32	75.5	5.1	1078	12	US-10-002-854-2	Sequence 2, Appli
33	75.5	5.1	3782	9	US-09-860-846-4	Sequence 4, Appli
34	75.5	5.1	3782	9	US-09-988-384B-4	Sequence 4, Appli
35	75.5	5.1	3782	9	US-09-836-821-4	Sequence 4, Appli
36	75.5	5.1	3782	10	US-09-861-289-4	Sequence 4, Appli
37	75	5.0	1385	10	US-09-738-363-2	Sequence 2, Appli
38	75	5.0	1864	9	US-09-832-292-27	Sequence 27, Appl
39	74.5	5.0	742	10	US-09-732-180-9	Sequence 9, Appli
40	74.5	5.0	2092	9	US-10-147-026-12	Sequence 12, Appl
41	74	5.0	517	10	US-09-964-277-21	Sequence 21, Appl
42	74	5.0	580	10	US-09-815-242-4959	Sequence 4959, Ap
43	74	5.0	589	10	US-09-815-242-10803	Sequence 10803, A
44	74	5.0	665	10	US-09-816-494-2	Sequence 2, Appli
45	74	5.0	665	10	US-09-964-277-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-729-674-279
; Sequence 279, Application US/09729674
; Patent No. US20010039335A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6055-64X
; CURRENT APPLICATION NUMBER: US/09/729,674
; CURRENT FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: 09/539,330
; PRIOR FILING DATE: 2000-03-30
; NUMBER OF SEQ ID NOS: 283
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 279
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-729-674-279

Query Match 10.1%; Score 151; DB 10; Length 147;
Best Local Similarity 35.0%; Pred. No. 1.7e-07;
Matches 41; Conservative 20; Mismatches 48; Indels 8; Gaps 2;

```
Qy      171 IPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLA-----NQKIGGDKK 225
            :|::|  :: | |::|||:|   |: ||  |||: ::   ||  ||     |   |
Db       23 LPRTLLDILADGTILKVGVGCSEDASKLLQDYGLVVRGCLDLRYLAMRQRNNLLCNG--- 79

Qy      226 WGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLP 282
```

```
Db         80 LSLKSLAETVLNFPLDKSLLLRCSNWDAETLTEDQVIYAARDAQISVALFLHLLGYP 136


RESULT 2
US-10-115-563-14
; Sequence 14, Application US/10115563
; Publication No. US20030008307A1
;   GENERAL INFORMATION:
;        APPLICANT: Griffin, John H
;                   Greengard, Judith S
;        TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
;                            C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
;                            AND COMPOSITIONS THEREOF
;        NUMBER OF SEQUENCES: 28
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: The Scripps Research Institute, Office of
;                        Patent Counsel
;             STREET: 10666 No. US20030008307A1th Torrey Pines Road, TPC 8
;             CITY: La Jolla
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 92037
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/115,563
;             FILING DATE: 02-Apr-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/410,488
;             FILING DATE: 24-MAR-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Fitting, Thomas
;             REGISTRATION NUMBER: 34,163
;             REFERENCE/DOCKET NUMBER: 449.0
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 619-554-2937
;             TELEFAX: 619-554-6312
;   INFORMATION FOR SEQ ID NO: 14:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2224 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-115-563-14

  Query Match             6.2%;  Score 93;  DB 9;  Length 2224;
  Best Local Similarity   19.2%;  Pred. No. 5.9;
  Matches   58;  Conservative   42;  Mismatches  124;  Indels   78;  Gaps   11;

Qy          4 SNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIR 63
Db        126 ASYLDHTFPAEKM--DDAVAPGREYTYEWSISEDSGPT---------HDDPP--CLTHIY 172

Qy         64 RQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRD 123
Db        173 YSHENLIEDFNSGLIGPLLICKK---------GTLTEGGTQKTFDKQIVLLFAVFDESK- 222

Qy        124 ESGIAFVGLDIEWRPSFR-----KGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHL 178
Db        223 -----------SWSQSSSLMYTVNGYVNGTMPDITVCA-----------HDHISWHLLGM 260

Qy        179 IEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLT------ 232
Db        261 SSGPEL--FSIHFNGQVLEQNHHKVSAITLVSATSTTANMTVGPEGKWIISSLTPKHLQA 318

Qy        233 ------------ETLVCKELLKPNRIRLGNWEFYPLSKQQL-QYAAT------DAYASW 272
Db        319 GMQAYIDIKNCPKKTRNLKKITREQRRHMKRWEYFIAAEEVIWDYAPVIPANMDKKYRSQ 378

Qy        273 HL 274
Db        379 HL 380


RESULT 3
US-09-815-242-11133
; Sequence 11133, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11133
;   LENGTH: 930
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-815-242-11133

  Query Match             6.0%;  Score 89.5;  DB 10;  Length 930;
  Best Local Similarity   20.7%;  Pred. No. 3.8;
  Matches   61;  Conservative   42;  Mismatches   93;  Indels   99;  Gaps   15;

Qy         24 ASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPRSITSSTSYKRFPLSR 83
Db        280 ARYEFKRWLNEVMNGADSITQTT------EQP---------------VKMNQYKATSQDQ 318

Qy         84 CRARNFPAMRFGGRILYSKTATEVD-------KRAMQLIKVLDTKRDE--------SGIA 128
Db        319 SAVENTPKIQI-DRTKYETLLTQADLTRWIEKLNAAKLIAV-DTETDSLDYMSANLVGIS 376

Qy        129 FVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQH---------LI 179
Db        377 F-------------ALENGEAAYLPLQLD--YLDA--------PKTLEKSTALAAIKPIL 413

Qy        180 EDSTLVKVGIGIDGDSVKLFHDYGVSIKDVE---------------DLSDLANQKIGGD 223
Db        414 ENPNIHKIGQNIKFDE-SIFARHGIELQGVEFDTMLLSYTLNSTGRHNMDDLAKRYLG-- 470

Qy        224 KKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVL 278
Db        471 --------HETIAFESLAGKGKSQL-TFNQIPL-EQATEYAAEDADVTMKLQQAL 515


RESULT 4
```

```
US-09-920-346-2
; Sequence 2, Application US/09920346
; Patent No. US20020081658A1
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A.J.
;  TITLE OF INVENTION: 18610, A NOVEL HUMAN TRANSIENT RECEPTOR
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: MNI-182
;  CURRENT APPLICATION NUMBER: US/09/920,346
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: USSN 60/221,925
;  PRIOR FILING DATE: 2000-07-31
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1885
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-920-346-2

  Query Match             5.9%;  Score 87.5;  DB 10;  Length 1885;
  Best Local Similarity   21.1%;  Pred. No. 17;
  Matches   54;  Conservative   47;  Mismatches   90;  Indels   65;  Gaps   12;

Qy          53 EDPNQIPNNIRR--QLPR---SITSSTSYKRFP----LSRCRARNFPAMRFGGRILYSKT 103
               |  | : | |    ::|:   : ||  : ||       : |   : |        ::||
Db        1464 EQQNDVRNVIMEYTEMPKYENNNTSENTLKRVSSLAGFTDCHRTSIP--------VHSKQ 1515

Qy         104 ATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFR-----KGVLPGKVATVQICVDS 158
               | :: :|        ||   :|  | :   :: ||| |     :| | |  :  :  :|:
Db        1516 AEKISRRP----STEDTHEVDSKAALIPDWLQDRPSNREMPSEEGTLNGLTSPFKPAMDT 1571

Qy         159 NYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGID-------GDSVKLFHDYGVSIKDVED 211
               ||       ::|          :| : |:::   |        |: | ::         :|:
Db        1572 NY------YYSA--------VERNNLMRLSQSIPFTPVPPRGEPVTVYR--------LEE 1609

Qy         212 LS-DLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYA 270
                | :: |  :    : || :  | |  :|:    |  :          | |  ::  |
Db        1610 SSPNILNNSMSSWSQLGLCAKIEFLSKEEMGGGLRRAV---------KVQCTWSEHDILK 1660

Qy         271 SWHLYKVLKDLPDAVS 286
               | ||| :   ||: |:
Db        1661 SGHLYIIKSFLPEVVN 1676


RESULT 5
US-10-174-363-22
; Sequence 22, Application US/10174363
; Publication No. US20030077623A1
; GENERAL INFORMATION:
;  APPLICANT: Butler, Karlene H.
;  APPLICANT:  Cahoon, Rebecca E
;  APPLICANT:  Harvell, Leslie T.
;  APPLICANT:  Rafalski, Antoni J.
;  APPLICANT:  Sakai, Hajime
;  TITLE OF INVENTION: Polynucleotides And Polypeptides Involved In Post-Transcriptional
;  TITLE OF INVENTION:  Gene Silencing
;  FILE REFERENCE: BB1454 US NA
;  CURRENT APPLICATION NUMBER: US/10/174,363
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: 60/298,973
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
;   LENGTH: 910
;   TYPE: PRT
;   ORGANISM: Zea mays
US-10-174-363-22

  Query Match             5.8%;  Score 87;  DB 9;  Length 910;
  Best Local Similarity   20.6%;  Pred. No. 6.7;
  Matches   66;  Conservative   40;  Mismatches   95;  Indels  120;  Gaps   16;

Qy           7 IDDAFTEEELLAIDA----------IEASYNFSRSSSSSSSAAPTVQATTSV-HGHEEDP 55
               ::| :    |: |:|          :|||     ||    |  ||:     | |||
Db         601 VNDPYLLNLLMKINAKLGGLNSLLQVEAS-----SSIPHVSQVPTIILGMDVSHGH---- 651

Qy          56 NQIPNNIRRQLPRSITSSTSYKRFPL-SRCRARNFPAMRFGGRI-----LYSKTATEVD- 108
                 |   |     |: :  | :::|| || ||          |:     |:    |: |
Db         652 ---PGQDR----PSVAAVVSSRQWPLISRYRA---SVHTQSARLEMMSSLFKPRGTDDDG 701

Qy         109 --------------KR------------AMQLIKVLDTKRDESGIAFVGLDIEWRPSFR 141
                           ||               |  :|:: : |:   |   || :| | |
Db         702 LIRESLIDFHTSSGKRKPEHIIIFRDGVSESQFTQVINIELDQIIEACKFLDEKWSPKFT 761

Qy         142 -----------------KGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQ----HLI 179
                              |||| |   ::|   |: |     |:|:  : :    |::
Db         762 VIVAQKNHHTKFFQTASPDNVLPGTVVDSKVCHPKNF-DFYMCAHAGMIGTTRPTHYHVL 820

Qy         180 EDSTLVKVGIGIDGDSVKLF----------------------HDYGVSIKD---VEDL 212
                |       ||   | :: |                      |     :     :|::
Db         821 HDE------IGFSADEMQEFVHSLSYVYQRSTTAISVVAPVCYAHLAAAQVSTFLRLEEM 874

Qy         213 SDLANQKIGGDKKWGLASLTE 233
               || :: : ||    | | : |
Db         875 SDASSSQGGGHTSAGSAPVPE 895


RESULT 6
US-09-741-669-380
; Sequence 380, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
;  APPLICANT:  Forsyth, R. Allyn
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  TITLE OF INVENTION:  Genes identified as required for
;  TITLE OF INVENTION:  proliferation of E. coli
;  FILE REFERENCE:  ELITRA.009A
;  CURRENT APPLICATION NUMBER: US/09/741,669
;  CURRENT FILING DATE:  2000-12-19
;  PRIOR APPLICATION NUMBER:   US 60/173005
;  PRIOR FILING DATE:   1999-12-23
;  NUMBER OF SEQ ID NOS:  481
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 380
;   LENGTH: 346
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-09-741-669-380

  Query Match             5.8%;  Score 86.5;  DB 10;  Length 346;
  Best Local Similarity   20.9%;  Pred. No. 1.9;
  Matches   54;  Conservative   34;  Mismatches   73;  Indels   97;  Gaps   12;

Qy          86 ARNFPAMRFGGRILYSKTATEV--------DKRAM----------QLIKVLDTKRDESG 126
               | | | :: |  :|  ||  ::        |  |:          | | |:  : | ||
Db         107 AVNIPGLKSGELVLDGKTLGDIYLGKIKKWDDEAIAKLNPGLKLPSQNIAVV-RRADGSG 165

Qy         127 IAFV------GLDIEWRPSFRKGV--------------------LPGKVATVQICVDS 158
                :||       :: ||: :   |                    ||| :  |:
Db         166 TSFVFTSYLAKVNEEWKNNVGTGSTVKWPIGLGGKGNDGIAAFVQRLPGAIGYVEYA--- 222

Qy         159 NYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLS----- 213
                |    :: :: :                |  ||  |    :   :     | |
Db         223 -YAKQNNLAYTKL----------------ISADGKPVSPTEENFANAAKGADWSKTFAQ 264

Qy         214 DLANQKIGGDKKWGLASLTETLVCKELLKPNR----IRLGNWEFYPLSKQ--QLQYAATD 267
               || |||  |:  | : | |  |: |:  || :    ::  :| :   :||   | ||:
Db         265 DLTNQK--GEDAWPITSTTFILIHKDQKKPEQGTEVLKFFDWAYKTGAKQANDLDYAS-- 320

Qy         268 AYASWHLYKVLKDLPDAV 285
                     |||:|
Db         321 -------------LPDSV 325
```

```
RESULT 7
US-10-268-441-15
; Sequence 15, Application US/10268441
; Publication No. US20030084475A1
; GENERAL INFORMATION:
;  APPLICANT: Cahoon, Edgar B.
;  APPLICANT:  Coughlan, Sean J.
;  APPLICANT:  Helentjaris, Timothy George
;  APPLICANT:  Jung, Rudolf
;  APPLICANT:  Li, Chun Ping
;  APPLICANT:  Nichols, Scott
;  APPLICANT:  Ripp, Kevin
;  APPLICANT:  Zheng, Peizhong
;  TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS AND PROTEINS AFFECTING STORAGE
;  TITLE OF INVENTION:  ORGANELLE
;  TITLE OF INVENTION:  FORMATION AND METHODS OF USE
;  FILE REFERENCE: BB1392 US NA
;  CURRENT APPLICATION NUMBER: US/10/268,441
;  CURRENT FILING DATE:  2002-10-09
;  PRIOR APPLICATION NUMBER: US/09/672,607
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: 60/157209
;  PRIOR FILING DATE: 1999-09-30
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
;   LENGTH: 367
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-10-268-441-15

  Query Match             5.8%;  Score 86.5;  DB 9;  Length 367;
  Best Local Similarity   31.1%;  Pred. No. 2;
  Matches   19;  Conservative   16;  Mismatches   25;  Indels    1;  Gaps    1;

Qy        176 QHLIEDSTLVKVGIGIDGDSVKLFHDYGVSI-KDVEDLSDLANQKIGGDKKWGLASLTET 234
              :|:||::   :  :|  |:   | :   ::| ||:| : ::| :   |||||    |||
Db         55 EHIIENAPTPRGHVGSSGNDPSLSYPLLIAILKDLETVVEVAAEHFYGDKKWNYIILTEA 114

Qy        235 L 235
              :
Db        115 M 115


RESULT 8
US-10-268-441-14
; Sequence 14, Application US/10268441
; Publication No. US20030084475A1
; GENERAL INFORMATION:
;  APPLICANT: Cahoon, Edgar B.
;  APPLICANT:  Coughlan, Sean J.
;  APPLICANT:  Helentjaris, Timothy George
;  APPLICANT:  Jung, Rudolf
;  APPLICANT:  Li, Chun Ping
;  APPLICANT:  Nichols, Scott
;  APPLICANT:  Ripp, Kevin
;  APPLICANT:  Zheng, Peizhong
;  TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS AND PROTEINS AFFECTING STORAGE
;  TITLE OF INVENTION:  ORGANELLE
;  TITLE OF INVENTION:  FORMATION AND METHODS OF USE
;  FILE REFERENCE: BB1392 US NA
;  CURRENT APPLICATION NUMBER: US/10/268,441
;  CURRENT FILING DATE:  2002-10-09
;  PRIOR APPLICATION NUMBER: US/09/672,607
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: 60/157209
;  PRIOR FILING DATE: 1999-09-30
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
;   LENGTH: 373
;   TYPE: PRT
;   ORGANISM: Triticum aestivum
US-10-268-441-14

  Query Match             5.8%;  Score 86.5;  DB 9;  Length 373;
  Best Local Similarity   32.1%;  Pred. No. 2.1;
  Matches   34;  Conservative   18;  Mismatches   31;  Indels   23;  Gaps    7;

Qy        144 VLPGKVATVQICVDSNYCDVMHIFHSGIPQSL-QHLIEDSTLVKVGIGIDGDSV-----K 197
              :|| : |  :|  :: |  :      ||  |: ||:||  |        || |:
Db         28 ILPERFANSEIAPEAVYALL------GIVSSVNQHIIETPT--------DGHSLASKEQS 73

Qy        198 LFHDYGVSI-KDVEDLSDLANQK-IGGDKKWGLASLTETL-VCKEL 240
              :     ||| |||| : ::| |  :| |:|||  ::|| :  |  |
Db         74 IPWALVVSILKDVEAVVEVAAQHFVGDDRKWGFLAVTEAVKACVRL 119


RESULT 9
US-09-815-242-12306
; Sequence 12306, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12306
;   LENGTH: 593
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12306

  Query Match             5.8%;  Score 86.5;  DB 10;  Length 593;
  Best Local Similarity   21.0%;  Pred. No. 4;
  Matches   55;  Conservative   47;  Mismatches  113;  Indels   47;  Gaps   11;

Qy          2 SSSNWIDDAFTEEELLAIDAI--EASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIP 59
              |  | |   ||  :  | |:  |:   :     : | | |  ||
Db        150 SYQNIISKVFTLPQDFTIIALTATATVEVQQDIREKLNIAQTDQIKTSTK---------- 199

Qy         60 NNIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLD 119
                || |   :  :   ::| |   :  :  |      |:|  |  :|:    :| : |:
Db        200 ---RRNLIFKVNPTYQRQKFILDYIKTHDEDA-----GIIYCSTRKQVE----ELQEALE 247

Qy        120 TKRDESGIAFVGLDIEWRPSFRKGVLPGK----VATVQICVDSNYCDVMHIFHSGIPQSL 175
              ::: || |   ||  : |   :   |  :    |||    :  :  :|  : |  :|  |
Db        248 SQKIESVIYHAGLSNKEREEAQNDFLFDRVKVVVATNAFGMGIDKSNVRFVIHYNMPGDL 307

Qy        176 QHLIEDSTLVKVG-IGIDGDSVKLF-------HDYGVSIKDVEDLSDLANQKIGGDKKWG 227
```

```
                :   :::   : |  |:  : : ||       |:| :::   :|  |  ::   |:|
Db        308 ESYYQEAG--RAGRDGLKSECILLFSERDINLHEYFITVSQADD--DYKDKM--GEKLTK 361

Qy        228 LASLTETLVCKE-----LLKPN 244
                :   |:|  | |       :||
Db        362 MIQYTKTKKCLEATIVHYFEPN 383


RESULT 10
US-09-888-615-74
; Sequence 74, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
;  APPLICANT: PLOWMAN, GREGORY
;  APPLICANT:  WHYTE, DAVID
;  APPLICANT:  CAENEPEEL, SEAN
;  APPLICANT:  CHARYDCZAK, GLEN
;  APPLICANT:  MANNING, GERARD
;  APPLICANT:  SUDARSANAM, SUCHA
;  TITLE OF INVENTION: NOVEL PROTEASES
;  FILE REFERENCE: 038602/1214
;  CURRENT APPLICATION NUMBER: US/09/888,615
;  CURRENT FILING DATE:  2001-06-26
;  PRIOR APPLICATION NUMBER: 60/214,047
;  PRIOR FILING DATE: 2000-06-26
;  NUMBER OF SEQ ID NOS: 150
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 74
;   LENGTH: 1042
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-888-615-74

  Query Match             5.8%;  Score 86;  DB 10;  Length 1042;
  Best Local Similarity   23.5%;  Pred. No. 10;
  Matches   72;  Conservative   42;  Mismatches  113;  Indels   80;  Gaps   15;

Qy          4 SNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIR 63
              || : : :  |    : :::    |:  ||: :|  |:|   : |          ||  :
Db        199 SNLLQNIWKAEPATLLPSLQEV--FASISSTDASFEPSVALASLV-------QHIPLQMI 249

Qy         64 RQLPRSITSSTSYKRFPLSR--CRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTK 121
               | ||:|:  : |   :::  ||  ::         |    |  ||   : |:|
Db        250 TVLIRSLTTDPNVKDASMTQALCRMIDW---------LSWPLAQHVDTWVIALLK----- 295

Qy        122 RDESGIAFVG----------LDIE------WRPSFRKG---VLPGKVATVQICVDSNYCD 162
                  |:| |           | ||      | |  | |   ||   : : |   :: :
Db        296 ----GLAAVQKFTILIDVTLLKIELVFNRLWFPLVRPGALAVLSHMLLSFQHSPEAFHLI 351

Qy        163 VMHI---FHS----GIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGV---------SI 206
              | |:    ||    |:| |   |:: : |:             ::|  |          :|
Db        352 VPHVVNLVHSFKNDGLPSSTAFLVQLTELI---------HCMMYHYSGFPDLYEPILEAI 402

Qy        207 KDV-----EDLSDLANQKIGGDKKWGLASLTETLVCK-ELLKPNRIRLGNWEFYPLSKQQ 260
              ||      | :  : ||      :   |||    |  | |  |   | |||   |  |  |
Db        403 KDFPKPSEEKIKLILNQSAWTSQSNSLASCLSRLSGKSETGKTGLINLGN-TCYMNSVIQ 461

Qy        261 LQYAATD 267
                : |||
Db        462 ALFMATD 468


RESULT 11
US-10-108-605-213
; Sequence 213, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
;  APPLICANT: Broadus, Julie
;  APPLICANT:  Stam, Lynn
;  APPLICANT:  Bachmann, Jane
;  APPLICANT:  Kamdar, Kim
;  TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
;  TITLE OF INVENTION:  PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
;  FILE REFERENCE: 31133B
;  CURRENT APPLICATION NUMBER: US/10/108,605
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: US 09/761,142
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/176,418
;  PRIOR FILING DATE: 2000-01-14
;  NUMBER OF SEQ ID NOS: 361
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 213
;   LENGTH: 1394
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-108-605-213

  Query Match             5.7%;  Score 85.5;  DB 9;  Length 1394;
  Best Local Similarity   22.4%;  Pred. No. 17;
  Matches   80;  Conservative   52;  Mismatches  140;  Indels   85;  Gaps   17;

Qy          8 DDAFTEEELLAIDAIEASYNFSR-----SSSSSSSAAPTVQATTSVHGHEEDPNQIPNNI 62
              ||   |:    :|: : :|   :      :: |:|:| | : | :|  :|:       :|
Db        371 DDELCEQHHQRLDSSQLNYLCQKFDEKLDTALSNSSANTGRNTPAVTANEDADGFFRRSI 430

Qy         63 RRQL---PRSITSSTSYKRFPLSRCR----------ARNFPAMRFG-------GRILYS- 101
              ::::   | :     |  |   :||:           :  |:|||        ||| :
Db        431 QQKIQYRPCTKNQQCSILRINRNRCQYCRLKKCIAVGMSRDAVRFGRVPKREKARILAAM 490

Qy        102 -----------KTATEVDKRAMQLIKVLD--------TKRDESGIAFVGLDIEWRPSFRK 142
                        |||:| :   |  ||          ||   | :     |    ||:
Db        491 QQSTQNRGQQRALATELDDQPRLLAAVLRAHLETCEFTKEKVSAMRQRARDC---PSYSM 547

Qy        143 GVL------PGKVATVQICVDSNYCDVMH--IFHSGIPQSLQHLIEDS--TLVKVGIGID 192
                 |      |     :     :  |:   |  :|:    | | :|   ||:| |:  |
Db        548 PTLLACPLNPAPELQSEQEFSQRFAHVIRGVIDFAGMIPGFQLLTQDDKFTLLKAGL-FD 606

Qy        193 GDSVKLFHDYGVSIKDV----------EDLSDLANQKIGGDKKWGLA------SLTET-- 234
                 |:|   :  ||  :          : : : || :   |  :  |      :||:
Db        607 ALFVRLICMFDSSINSIICLNGQVMRRDAIQNGANARFLVDSTFNFAERMNSMNLTDAEI 666

Qy        235 -LVCK-ELLKPNRIRLGNWE----FYPLSKQQLQY--AATDAYASWHLYKVLKDLPD 283
               | |   |: |:|  | | |      |   |  |||  |          | |:|: :||
Db        667 GLFCAIVLITPDRPGLRNLELIEKMYSRLKGCLQYIVAQNRPDQPEFLAKLLETMPD 723


RESULT 12
US-09-738-626-6900
; Sequence 6900, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6900
;   LENGTH: 319
```

```
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-6900

  Query Match             5.6%;  Score 83.5;  DB 9;  Length 319;
  Best Local Similarity   17.0%;  Pred. No. 3.4;
  Matches   52;  Conservative   46;  Mismatches   88;  Indels  119;  Gaps   13;

Qy         31 SSSSSSSAA------PTVQATTSVHGHEEDPNQIPNNIRRQLPRSITSSTSYKRFPLSRC 84
              ::|:|   |       |  : |:  | : :| | |
Db         13 ATSTSDDVAVGGRLFSTADSATAALGSDAEPGQFPRT---------------------- 49

Qy         85 RARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDES----GIAFVGL----DIEW 136
                     :::|:  | ::::  ::: |||:   :     |:  ||:    |
Db         50 ------------VVHSRGETTLEQQPQRVV-VLDSGEIDQVLSLGVTPVGIASPKDASS 95

Qy        137 RPSFRKGVL--------------------PGKVATVQICVDSNYCDVMHI------FHS 169
              :|:: :  |                    |  :   :: || :|  :  |
Db         96 QPAYLENQLADVQTVGTTSELNFEAIAALKPDLILGSKLRVDESYDQLSQIAPTVLSIRP 155

Qy        170 GIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLA 229
              | |       |:  |    :|::| :|:: ::|   :       |   : | |  :  |
Db        156 GFPWK-----ENFLLTADALGLEGKAVEVLNEYQTHV-------DAVRETIDGSPEISLV 203

Qy        230 SLTETLVCKELLKPNRIRL-GNWEF-------YPLSKQQLQYA-----------ATDAYA 270
                  | | || ||  |          ||: ::|                   |||
Db        204 R----------FMPGRTRLYGNLSFIGAILKDLGLSRPEIQNIDDLAVEISPENITDANG 253

Qy        271 SWHLY 275
               |  |
Db        254 DWIFY 258


RESULT 13
US-09-738-626-6005
; Sequence 6005, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6005
;   LENGTH: 305
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-6005

  Query Match             5.5%;  Score 82.5;  DB 9;  Length 305;
  Best Local Similarity   21.9%;  Pred. No. 4;
  Matches   41;  Conservative   29;  Mismatches   62;  Indels   55;  Gaps    7;

Qy        127 IAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLI------E 180
              ::|||     ||:  |  |   :   :| : :|            |:: :| |      :
Db         15 VSFVG-----RPNTGKSTLTNALVGEKIAITANQ----------PETTRHPIRGLVHRD 58

Qy        181 DSTLVKVGI-GIDGDSVKLFHDYGVSIKDVEDLSDL------ANQKIGGDKKWGLA---- 229
              :: :: |   |:       |       ::||      ||       ||:|||   :| |
Db         59 NAQIIVVDTPGLHRPRTLLGERLNEAVKDTYADVDLIGFTVPANEKIGPGDRWILEAVRK 118

Qy        230 --------------SLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHL 274
                           |::  ||  :|:  : :  || |  |:|    :   |       |
Db        119 VSPKTPILGIITKADSVSRDLVAAQLMAVHELLGGNSEVVPVSSTSGENVET-------L 171

Qy        275 YKVLKDL 281
                ||: ||
Db        172 IKVMTDL 178


RESULT 14
US-10-174-363-50
; Sequence 50, Application US/10174363
; Publication No. US20030077623A1
; GENERAL INFORMATION:
;  APPLICANT: Butler, Karlene H.
;  APPLICANT:  Cahoon, Rebecca E
;  APPLICANT:  Harvell, Leslie T.
;  APPLICANT:  Rafalski, Antoni J.
;  APPLICANT:  Sakai, Hajime
;  TITLE OF INVENTION: Polynucleotides And Polypeptides Involved In Post-Transcriptio
;  TITLE OF INVENTION:  Gene Silencing
;  FILE REFERENCE: BB1454 US NA
;  CURRENT APPLICATION NUMBER: US/10/174,363
;  CURRENT FILING DATE:  2002-06-17
;  PRIOR APPLICATION NUMBER: 60/298,973
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 50
;   LENGTH: 389
;   TYPE: PRT
;   ORGANISM: Triticum aestivum
US-10-174-363-50

  Query Match             5.4%;  Score 81;  DB 9;  Length 389;
  Best Local Similarity   21.1%;  Pred. No. 8;
  Matches   36;  Conservative   24;  Mismatches   53;  Indels   58;  Gaps    6;

Qy        113 QLIKVLDTKRDESGIAFVGLDIEWRPSFR-----------------KGVLPGKVATVQI 154
              |  :|:: : ::   |   || :| | |                  : | || |   |:
Db        212 QFTQVINIELEQIIEACKCLDDKWEPKFTVIVAQKNHHTRFFQTNSPENVPPGTVVDKQV 271

Qy        155 CVDSNYCDVMHIFHSGIPQSLQ----HLIEDSTLVKVGIGIDGDSVKLF---------- 199
              |   |: |     |:|:  : :    |:: |       ||  || :: |
Db        272 CHPKNF-DFYMCAHAGMIGTSRPTHYHVLHDE------IGFSGDELQEFVHSLSYVYQRS 324

Qy        200 -----------------HDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTE 233
                              |  :|  ||:|| :: : ||    | | : |
Db        325 TTAISVAAPIAYAHLAAAQVGTFMK-FEDMSDTSSSQGGGHTSAGSAPVPE 374


RESULT 15
US-09-391-340-12
; Sequence 12, Application US/09391340A
; Patent No. US20020013455A1
; GENERAL INFORMATION:
;  APPLICANT: Callen, Walter
;  APPLICANT: Mather, Eric
;  TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
;  FILE REFERENCE: 09010/027001
;  CURRENT APPLICATION NUMBER: US/09/391,340A
;  CURRENT FILING DATE: 1999-09-07
;  EARLIER APPLICATION NUMBER: US 08/907,166
;  EARLIER FILING DATE: 1997-08-06
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12
;   LENGTH: 574
```

```
;   TYPE: PRT
;   ORGANISM: Aquifex pyrophilus
US-09-391-340-12

  Query Match             5.4%;  Score 80;  DB 10;  Length 574;
  Best Local Similarity   21.9%;  Pred. No. 18;
  Matches   33;  Conservative   30;  Mismatches   48;  Indels   40;  Gaps    6;

Qy        170 GIPQSLQHLIEDSTLVKVGIGIDGDSVKLF------HDYGV---SIKDVEDLSDLANQK- 219
              |: :::: | |:|  : :     || ::|       : | :    |:|:| |  | |::
Db         11 GLKKAIKRL-ENSPYLYLDTETTGDRIRLVQIGDEENTYVIDLYEIQDIEPLRKLINERG 69

Qy        220 -IGGDKKWGL----------ASLTETLVCKELLKPNRIRL-----------------GN 250
               :| : |: |          ::  :|::   ||   | |                  :
Db         70 IVGHNLKFDLKYLYRYGIFPSATFDTMIASYLLGYERHSLNHIVSNLLGYSMDKSYQTSD 129

Qy        251 WEFYPLSKQQLQYAATDAYASWHLYKVLKDL 281
              |    ||  ||:||| |      |:  ::|:
Db        130 WGASVLSDAQLKYAANDVIVLRELFPKMRDM 160


Search completed: June  6, 2003, 10:31:31
Job time : 23 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 6, 2003, 10:22:55 ; Search time 308 Seconds
(without alignments)
602.867 Million cell updates/sec

Title: US-09-896-186B-24
Perfect score: 1491
Sequence: 1 MSSSNWIDDAFTEEELLAID.....YASWHLYKVLKDLPDAVSGS 288

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US080_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/1/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/1/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/1/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/1/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/1/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/1/paa/US102_COMB.pep:*
27: /cgn2_6/ptodata/1/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1491	100.0	288	19	US-09-513-996A-67237	Sequence 67237, A
2	1491	100.0	288	22	US-09-896-186B-24	Sequence 24, Appl
3	1491	100.0	298	19	US-09-513-996A-67236	Sequence 67236, A
4	1439.5	96.5	313	21	US-09-708-427-27138	Sequence 27138, A
5	1439.5	96.5	313	22	US-09-896-186B-2	Sequence 2, Appli
6	1439.5	96.5	313	23	US-09-906-226-50	Sequence 50, Appl
7	1433	96.1	288	19	US-09-513-996A-28856	Sequence 28856, A
8	1028	68.9	197	19	US-09-513-996A-67238	Sequence 67238, A
9	996	66.8	197	19	US-09-513-996A-28857	Sequence 28857, A
10	976.5	65.5	222	21	US-09-708-427-27139	Sequence 27139, A
11	896	60.1	177	19	US-09-513-996A-28858	Sequence 28858, A
12	876.5	58.8	202	21	US-09-708-427-27140	Sequence 27140, A
13	717	48.1	157	20	US-09-620-393B-6113	Sequence 6113, Ap
14	676.5	45.4	290	23	US-09-906-226-16	Sequence 16, Appl
15	676.5	45.4	290	27	US-60-218-993-16	Sequence 16, Appl
16	545.5	36.6	301	21	US-09-708-427-65911	Sequence 65911, A
17	545.5	36.6	301	21	US-09-708-427-83273	Sequence 83273, A
18	545.5	36.6	303	21	US-09-708-427-65910	Sequence 65910, A
19	545.5	36.6	303	21	US-09-708-427-83272	Sequence 83272, A
20	475	31.9	173	21	US-09-708-427-65912	Sequence 65912, A
21	475	31.9	173	21	US-09-708-427-83274	Sequence 83274, A
22	352	23.6	1409	23	US-09-948-941-397	Sequence 397, App
23	352	23.6	1432	9	US-08-594-242-71	Sequence 71, Appl
24	352	23.6	1432	20	US-09-618-166-71	Sequence 71, Appl
25	352	23.6	1432	21	US-09-791-537-602	Sequence 602, App
26	352	23.6	1432	22	US-09-896-186B-18	Sequence 18, Appl
27	352	23.6	1432	23	US-09-948-941-273	Sequence 273, App
28	339.5	22.8	1436	21	US-09-791-537-106290	Sequence 106290,
29	330.5	22.2	643	21	US-09-791-537-82396	Sequence 82396, A
30	330.5	22.2	643	23	US-09-906-226-49	Sequence 49, Appl
31	330	22.1	1401	20	US-09-618-166-206	Sequence 206, App
32	330	22.1	1401	21	US-09-791-537-128052	Sequence 128052,
33	270	18.1	216	23	US-09-906-226-38	Sequence 38, Appl
34	254	17.0	234	23	US-09-906-226-14	Sequence 14, Appl
35	254	17.0	234	27	US-60-218-993-14	Sequence 14, Appl
36	249.5	16.7	210	26	US-10-219-999-33507	Sequence 33507, A
37	249.5	16.7	210	27	US-60-324-109-18761	Sequence 18761, A
38	249.5	16.7	213	23	US-09-906-226-10	Sequence 10, Appl
39	249.5	16.7	213	27	US-60-218-993-10	Sequence 10, Appl
40	245	16.4	236	26	US-10-219-999-60555	Sequence 60555, A
41	242.5	16.3	241	26	US-10-219-999-56100	Sequence 56100, A
42	238	16.0	197	23	US-09-906-226-12	Sequence 12, Appl
43	238	16.0	197	27	US-60-218-993-12	Sequence 12, Appl
44	238	16.0	307	27	US-60-243-468-753	Sequence 753, App
45	237.5	15.9	230	26	US-10-219-999-48774	Sequence 48774, A

ALIGNMENTS

RESULT 1
US-09-513-996A-67237
; Sequence 67237, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 67237
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..288
; OTHER INFORMATION: any n or Xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..288 / Ceres Seq. ID 2177599
US-09-513-996A-67237

Query Match 100.0%; Score 1491; DB 19; Length 288;
Best Local Similarity 100.0%; Pred. No. 6.7e-154;
Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

```
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288


RESULT 2
US-09-896-186B-24
; Sequence 24, Application US/09896186B
; GENERAL INFORMATION:
;  APPLICANT: Joshua Z. Levin
;  APPLICANT:  Ken Phillips
;  APPLICANT:  Greg Budziszewski
;  APPLICANT:  Fred Meins
;  APPLICANT:  Zhenya Glazov
;  TITLE OF INVENTION: Methods of Controlling Gene Expression
;  FILE REFERENCE: PB/5-31481A
;  CURRENT APPLICATION NUMBER: US/09/896,186B
;  CURRENT FILING DATE:  2002-04-04
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 288
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-896-186B-24

  Query Match            100.0%;  Score 1491;  DB 22;  Length 288;
  Best Local Similarity  100.0%;  Pred. No. 6.7e-154;
  Matches  288;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288


RESULT 3
US-09-513-996A-67236
; Sequence 67236, Application US/09513996A
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
;  TITLE OF INVENTION:  ENCODED THEREBY
;  FILE REFERENCE: 2750-709P
;  CURRENT APPLICATION NUMBER: US/09/513,996A
;  CURRENT FILING DATE:  2000-02-25
;  NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 67236
;   LENGTH: 298
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: 1..298
;   OTHER INFORMATION: any n or Xaa = unknown
;   FEATURE:
;   OTHER INFORMATION: Location 1..298 / Ceres Seq. ID 2177598
US-09-513-996A-67236

  Query Match            100.0%;  Score 1491;  DB 19;  Length 298;
  Best Local Similarity  100.0%;  Pred. No. 7.1e-154;
  Matches  288;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 70

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 130

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 190

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 250

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 298


RESULT 4
US-09-708-427-27138
; Sequence 27138, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27138
;   LENGTH: 313
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..313
;   OTHER INFORMATION: Xaa is any amino acid
;   NAME/KEY: misc_feature
;   LOCATION: 1..313
;   OTHER INFORMATION: Ceres Seq. ID 1820253
US-09-708-427-27138

  Query Match            96.5%;  Score 1439.5;  DB 21;  Length 313;
  Best Local Similarity  95.3%;  Pred. No. 3.4e-148;
  Matches  281;  Conservative    1;  Mismatches    4;  Indels    9;  Gaps    1;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
```

```
Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKV---------LKDLPDAVS 286
              ||||||||||||||||||||||||||||||||||||         | ||   :|
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVTTTKNHLLTLNDLEAKIS 295


RESULT 5
US-09-896-186B-2
; Sequence 2, Application US/09896186B
; GENERAL INFORMATION:
;  APPLICANT: Joshua Z. Levin
;  APPLICANT:  Ken Phillips
;  APPLICANT:  Greg Budziszewski
;  APPLICANT:  Fred Meins
;  APPLICANT:  Zhenya Glazov
;  TITLE OF INVENTION: Methods of Controlling Gene Expression
;  FILE REFERENCE: PB/5-31481A
;  CURRENT APPLICATION NUMBER: US/09/896,186B
;  CURRENT FILING DATE:  2002-04-04
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 313
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-896-186B-2

  Query Match             96.5%;  Score 1439.5;  DB 22;  Length 313;
  Best Local Similarity   95.3%;  Pred. No. 3.4e-148;
  Matches  281;  Conservative    1;  Mismatches    4;  Indels    9;  Gaps    1;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKV---------LKDLPDAVS 286
              ||||||||||||||||||||||||||||||||||||         | ||   :|
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVTTTKNHLLTLNDLEAKIS 295


RESULT 6
US-09-906-226-50
; Sequence 50, Application US/09906226
; GENERAL INFORMATION:
;  APPLICANT: Butler, Karla
;  APPLICANT:  Cahoon, Rebecca E.
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Sakai, Hajime
;  TITLE OF INVENTION: Plant RNaseD-Like Genes
;  FILE REFERENCE: BB1467 US NA
;  CURRENT APPLICATION NUMBER: US/09/906,226
;  CURRENT FILING DATE:  2001-07-16
;  PRIOR APPLICATION NUMBER: 60/218993
;  PRIOR FILING DATE: July 17, 2000
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 50
;   LENGTH: 313
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-906-226-50

  Query Match             96.5%;  Score 1439.5;  DB 23;  Length 313;
  Best Local Similarity   95.3%;  Pred. No. 3.4e-148;
  Matches  281;  Conservative    1;  Mismatches    4;  Indels    9;  Gaps    1;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKV---------LKDLPDAVS 286
              ||||||||||||||||||||||||||||||||||||         | ||   :|
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVTTTKNHLLTLNDLEAKIS 295


RESULT 7
US-09-513-996A-28856
; Sequence 28856, Application US/09513996A
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  ENCODED THEREBY
;  FILE REFERENCE: 2750-709P
;  CURRENT APPLICATION NUMBER: US/09/513,996A
;  CURRENT FILING DATE:  2000-02-25
;  NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 28856
;   LENGTH: 288
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: 1..288
;   OTHER INFORMATION: any n or Xaa = unknown
;   FEATURE:
;   OTHER INFORMATION: Location 1..288 / Ceres Seq. ID 1572652
US-09-513-996A-28856

  Query Match             96.1%;  Score 1433;  DB 19;  Length 288;
  Best Local Similarity   96.5%;  Pred. No. 1.5e-147;
  Matches  278;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSLSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              |||||||||||  || ||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITCPTSSKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIE 180
```

```
              |||||| |||| |||||||||||||||||||||||||:|||||||||||||||||||||
Db        121 KRDESGRAFVGFDIEWRPSFRKGVLPGKVATVQICVDNNYCDVMHIFHSGIPQSLQHLIE 180

Qy        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKEL 240
              |||||||||||||||||||||||||||||||||||||| |:|||||||||||||||||
Db        181 DSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIDGEKKWGLASLTETLVCKEL 240

Qy        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
              |||||||||||||||||||||||||||||||||||||||||||||| |
Db        241 LKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSDS 288


RESULT 8
US-09-513-996A-67238
; Sequence 67238, Application US/09513996A
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
;  TITLE OF INVENTION:  ENCODED THEREBY
;  FILE REFERENCE: 2750-709P
;  CURRENT APPLICATION NUMBER: US/09/513,996A
;  CURRENT FILING DATE:  2000-02-25
;  NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 67238
;   LENGTH: 197
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: 1..197
;   OTHER INFORMATION: any n or Xaa = unknown
;   FEATURE:
;   OTHER INFORMATION: Location 1..197 / Ceres Seq. ID 2177600
US-09-513-996A-67238

  Query Match             68.9%;  Score 1028;  DB 19;  Length 197;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-103;
  Matches  197;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVAT 151
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVAT 60

Qy        152 VQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 211
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 120

Qy        212 LSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYAS 271
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYAS 180

Qy        272 WHLYKVLKDLPDAVSGS 288
              |||||||||||||||||
Db        181 WHLYKVLKDLPDAVSGS 197


RESULT 9
US-09-513-996A-28857
; Sequence 28857, Application US/09513996A
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
;  TITLE OF INVENTION:  ENCODED THEREBY
;  FILE REFERENCE: 2750-709P
;  CURRENT APPLICATION NUMBER: US/09/513,996A
;  CURRENT FILING DATE:  2000-02-25
;  NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 28857
;   LENGTH: 197
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: 1..197
;   OTHER INFORMATION: any n or Xaa = unknown
;   FEATURE:
;   OTHER INFORMATION: Location 1..197 / Ceres Seq. ID 1572653
US-09-513-996A-28857

  Query Match             66.8%;  Score 996;  DB 19;  Length 197;
  Best Local Similarity   97.0%;  Pred. No. 6.9e-100;
  Matches  191;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         92 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVAT 151
              |||||||||||||||||||||||||||||||| |||| ||||||||||||||||||
Db          1 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGRAFVGFDIEWRPSFRKGVLPGKVAT 60

Qy        152 VQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 211
              ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VQICVDNNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 120

Qy        212 LSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYAS 271
              ||||||||| |:|||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LSDLANQKIDGEKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYAS 180

Qy        272 WHLYKVLKDLPDAVSGS 288
              ||||||||||||||| |
Db        181 WHLYKVLKDLPDAVSDS 197


RESULT 10
US-09-708-427-27139
; Sequence 27139, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27139
;   LENGTH: 222
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..222
;   OTHER INFORMATION: Xaa is any amino acid
;   NAME/KEY: misc_feature
;   LOCATION: 1..222
;   OTHER INFORMATION: Ceres Seq. ID 1820254
US-09-708-427-27139

  Query Match             65.5%;  Score 976.5;  DB 21;  Length 222;
  Best Local Similarity   93.1%;  Pred. No. 1.2e-97;
  Matches  190;  Conservative    1;  Mismatches    4;  Indels    9;  Gaps    1;

Qy         92 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVAT 151
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVAT 60

Qy        152 VQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVED 120

Qy        212 LSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYAS 271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYAS 180

Qy        272 WHLYKV---------LKDLPDAVS 286
              ||||||         | ||   :|
Db        181 WHLYKVTTTKNHLLTLNDLEAKIS 204
```

```
RESULT 11
US-09-513-996A-28858
; Sequence 28858, Application US/09513996A
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
;  TITLE OF INVENTION:  ENCODED THEREBY
;  FILE REFERENCE: 2750-709P
;  CURRENT APPLICATION NUMBER: US/09/513,996A
;  CURRENT FILING DATE:  2000-02-25
;  NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 28858
;   LENGTH: 177
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: 1..177
;   OTHER INFORMATION: any n or Xaa = unknown
;   FEATURE:
;   OTHER INFORMATION: Location 1..177 / Ceres Seq. ID 1572654
US-09-513-996A-28858

  Query Match             60.1%;  Score 896;  DB 19;  Length 177;
  Best Local Similarity   96.6%;  Pred. No. 5.4e-89;
  Matches  171;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        112 MQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGI 171
              ||||||||||||||| |||| |||||||||||||||||||:||||||||||||
Db          1 MQLIKVLDTKRDESGRAFVGFDIEWRPSFRKGVLPGKVATVQICVDNNYCDVMHIFHSGI 60

Qy        172 PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASL 231
              |||||||||||||||||||||||||||||||||||||||||||||||| |:||||||||
Db         61 PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIDGEKKWGLASL 120

Qy        232 TETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSGS 288
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        121 TETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDAVSDS 177


RESULT 12
US-09-708-427-27140
; Sequence 27140, Application US/09708427
; GENERAL INFORMATION:
;  APPLICANT: N. ALEXANDROV et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1243P
;  CURRENT APPLICATION NUMBER: US/09/708,427
;  CURRENT FILING DATE:  2000-11-09
;  NUMBER OF SEQ ID NOS: 85364
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27140
;   LENGTH: 202
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..202
;   OTHER INFORMATION: Xaa is any amino acid
;   NAME/KEY: misc_feature
;   LOCATION: 1..202
;   OTHER INFORMATION: Ceres Seq. ID 1820255
US-09-708-427-27140

  Query Match             58.8%;  Score 876.5;  DB 21;  Length 202;
  Best Local Similarity   92.4%;  Pred. No. 9.1e-87;
  Matches  170;  Conservative    1;  Mismatches    4;  Indels    9;  Gaps    1;

Qy        112 MQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGI 171
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGI 60

Qy        172 PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASL 231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASL 120

Qy        232 TETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKV---------LKDLP 282
              ||||||||||||||||||||||||||||||||||||||||||||||         | ||
Db        121 TETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVTTTKNHLLTLNDLE 180

Qy        283 DAVS 286
                :|
Db        181 AKIS 184


RESULT 13
US-09-620-393B-6113
; Sequence 6113, Application US/09620393B
; GENERAL INFORMATION:
;  APPLICANT: ALEXANDROV, Nickolai et al.
;  TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
;  TITLE OF INVENTION:  THEREBY
;  FILE REFERENCE: 2750-1068P
;  CURRENT APPLICATION NUMBER: US/09/620,393B
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 9948
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6113
;   LENGTH: 157
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..157
;   OTHER INFORMATION: Xaa is any amino acid
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..157
;   OTHER INFORMATION: Ceres Seq. ID 1392053
US-09-620-393B-6113

  Query Match             48.1%;  Score 717;  DB 20;  Length 157;
  Best Local Similarity   95.9%;  Pred. No. 1.9e-69;
  Matches  142;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 MSSSNWIDDAFTEEELLAIDAIEASYNFSLSSSSSSSAAPTVQATTSVHGHEEDPNQIPN 60

Qy         61 NIRRQLPRSITSSTSYKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120
              ||||||||||  || ||||||||||||||||||||||||||||||||||||||||||||
Db         61 NIRRQLPRSITCPTSSKRFPLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDT 120

Qy        121 KRDESGIAFVGLDIEWRPSFRKGVLPGK 148
              |||||| |||| ||||||||||||||||
Db        121 KRDESGRAFVGFDIEWRPSFRKGVLPGK 148


RESULT 14
US-09-906-226-16
; Sequence 16, Application US/09906226
; GENERAL INFORMATION:
;  APPLICANT: Butler, Karla
;  APPLICANT:  Cahoon, Rebecca E.
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Sakai, Hajime
;  TITLE OF INVENTION: Plant RNaseD-Like Genes
;  FILE REFERENCE: BB1467 US NA
;  CURRENT APPLICATION NUMBER: US/09/906,226
;  CURRENT FILING DATE:  2001-07-16
;  PRIOR APPLICATION NUMBER: 60/218993
;  PRIOR FILING DATE: July 17, 2000
;  NUMBER OF SEQ ID NOS: 57
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
```

```
;   LENGTH: 290
;   TYPE: PRT
;   ORGANISM: Glycine max
US-09-906-226-16

  Query Match           45.4%;  Score 676.5;  DB 23;  Length 290;
  Best Local Similarity   49.3%;  Pred. No. 1.4e-64;
  Matches  139;  Conservative   49;  Mismatches   63;  Indels   31;  Gaps    6;

Qy          9 DAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPR 68
              :  ||:::| ||:|          | |::   |        : |   |       ||:||:
Db         18 EPLTEDDLEAIEA----------SLSNNKKRP-------FNDHTHTP-------RRRLPK 53

Qy         69 SI-----TSSTSYKRFPLSRCRAR-NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKR 122
              |:      :::|:   |   | :|   | |:| |:| ||:|   |:| | :|:::|  |
Db         54 SLIALQHPNASSFSPHP-RPCDSRMTLPVMKFSGQISYSRTFDAVEKAATKLLQILQEKT 112

Qy        123 DESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIEDS 182
               :     :| ||||:|:||||| ||||| :||| |: :| |:|: ||||||:|| |:||
Db        113 TDMMQTAIGFDIEWKPTFRKGVPPGKVAVMQICGDTRHCHVLHLIHSGIPQNLQLLLEDP 172

Qy        183 TLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLK 242
              |::||| |||||:||:| || :|:| | |||  ||||:||| |||||||||| |: |:| |
Db        173 TVLKVGAGIDGDAVKVFRDYNISVKGVTDLSFHANQKLGGDHKWGLASLTEKLLSKQLKK 232

Qy        243 PNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDA 284
              ||:|||||||   |||:||:||||||:||| ||: :||||||
Db        233 PNKIRLGNWEAPVLSKEQLEYAATDAFASWCLYQAIKDLPDA 274


RESULT 15
US-60-218-993-16
; Sequence 16, Application US/60218993
; GENERAL INFORMATION:
;  APPLICANT:  Butler, Karla
;  APPLICANT:  Cahoon, Rebecca E.
;  APPLICANT:  Rafalski, Antoni
;  TITLE OF INVENTION:  Plant RNaseD-Like Genes
;  FILE REFERENCE:  BB1467 US PRV
;  CURRENT APPLICATION NUMBER: US/60/218,993
;  CURRENT FILING DATE:  2000-07-17
;  NUMBER OF SEQ ID NOS:  22
;  SOFTWARE:  Microsoft Office 97
; SEQ ID NO 16
;   LENGTH: 290
;   TYPE: PRT
;   ORGANISM: Glycine max
US-60-218-993-16

  Query Match           45.4%;  Score 676.5;  DB 27;  Length 290;
  Best Local Similarity   49.3%;  Pred. No. 1.4e-64;
  Matches  139;  Conservative   49;  Mismatches   63;  Indels   31;  Gaps    6;

Qy          9 DAFTEEELLAIDAIEASYNFSRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPR 68
              :  ||:::| ||:|          | |::   |        : |   |       ||:||:
Db         18 EPLTEDDLEAIEA----------SLSNNKKRP-------FNDHTHTP-------RRRLPK 53

Qy         69 SI-----TSSTSYKRFPLSRCRAR-NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKR 122
              |:      :::|:   |   | :|   | |:| |:| ||:|   |:| | :|:::|  |
Db         54 SLIALQHPNASSFSPHP-RPCDSRMTLPVMKFSGQISYSRTFDAVEKAATKLLQILQEKT 112

Qy        123 DESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIEDS 182
               :     :| ||||:|:||||| ||||| :||| |: :| |:|: ||||||:|| |:||
Db        113 TDMMQTAIGFDIEWKPTFRKGVPPGKVAVMQICGDTRHCHVLHLIHSGIPQNLQLLLEDP 172

Qy        183 TLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLK 242
              |::||| |||||:||:| || :|:| | |||  ||||:||| |||||||||| |: |:| |
Db        173 TVLKVGAGIDGDAVKVFRDYNISVKGVTDLSFHANQKLGGDHKWGLASLTEKLLSKQLKK 232

Qy        243 PNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLPDA 284
              ||:|||||||   |||:||:||||||:||| ||: :||||||
Db        233 PNKIRLGNWEAPVLSKEQLEYAATDAFASWCLYQAIKDLPDA 274
```

```
Search completed: June  6, 2003, 10:30:14
Job time : 309 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 6, 2003, 10:23:40 ; Search time 41 Seconds
(without alignments)
1513.077 Million cell updates/sec

Title: US-09-896-186B-24
Perfect score: 1491
Sequence: 1 MSSSNWIDDAFTEEELLAID.....YASWHLYKVLKDLPDAVSGS 288

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1023978 seqs, 215403328 residues

Total number of hits satisfying chosen parameters: 1023978

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	666.5	44.7	268	6	US-10-424-599-227379	Sequence 227379,
2	352	23.6	994	5	US-09-724-676-89006	Sequence 89006, A
3	352	23.6	994	5	US-09-724-676A-89006	Sequence 89006, A
4	352	23.6	1409	5	US-09-949-001-22	Sequence 22, Appl
5	352	23.6	1432	5	US-09-949-001-16	Sequence 16, Appl
6	352	23.6	1432	6	US-10-374-077-71	Sequence 71, Appl
7	330	22.1	1401	6	US-10-374-077-206	Sequence 206, App
8	253.5	17.0	238	6	US-10-424-599-238476	Sequence 238476,
9	249.5	16.7	210	6	US-10-425-114-37706	Sequence 37706, A
10	245	16.4	236	6	US-10-425-114-71107	Sequence 71107, A
11	212	14.2	208	6	US-10-424-599-191095	Sequence 191095,
12	188	12.6	205	6	US-10-424-599-229443	Sequence 229443,
13	169.5	11.4	223	6	US-10-431-652-6289	Sequence 6289, Ap
14	163.5	11.0	198	6	US-10-425-114-71315	Sequence 71315, A
15	151	10.1	127	6	US-10-276-774-1634	Sequence 1634, Ap
16	142	9.5	599	6	US-10-094-749-1828	Sequence 1828, Ap
17	119.5	8.0	202	6	US-10-424-599-254772	Sequence 254772,
18	117	7.8	511	6	US-10-425-114-43332	Sequence 43332, A
19	99	6.6	126	6	US-10-424-599-220783	Sequence 220783,
20	94.5	6.3	159	6	US-10-424-599-274357	Sequence 274357,
21	93	6.2	488	6	US-10-369-493-18241	Sequence 18241, A
22	93	6.2	1348	5	US-09-949-002-517	Sequence 517, App
23	93	6.2	2207	5	US-09-724-676-60068	Sequence 60068, A
24	93	6.2	2207	5	US-09-724-676A-60068	Sequence 60068, A
25	93	6.2	2224	1	PCT-US02-19017-31	Sequence 31, Appl
26	93	6.2	2224	5	US-09-949-002-292	Sequence 292, App
27	93	6.2	2224	7	US-60-452-680-22968	Sequence 22968, A
28	93	6.2	2224	7	US-60-453-135-14275	Sequence 14275, A
29	93	6.2	2224	7	US-60-453-050-14275	Sequence 14275, A
30	93	6.2	2224	7	US-60-455-444-7663	Sequence 7663, Ap
31	93	6.2	2224	7	US-60-465-241-7663	Sequence 7663, Ap
32	93	6.2	2224	7	US-60-466-412-14275	Sequence 14275, A
33	93	6.2	3079	6	US-10-369-493-2024	Sequence 2024, Ap
34	92.5	6.2	481	6	US-10-431-652-4475	Sequence 4475, Ap
35	91.5	6.1	1836	5	US-09-949-016-7432	Sequence 7432, Ap
36	89.5	6.0	357	6	US-10-417-886-8398	Sequence 8398, Ap
37	89.5	6.0	930	6	US-10-282-122A-58327	Sequence 58327, A
38	89.5	6.0	1324	1	PCT-US02-32637-53	Sequence 53, Appl
39	89.5	6.0	1324	6	US-10-270-839-53	Sequence 53, Appl
40	89.5	6.0	1324	7	US-60-427-165-32	Sequence 32, Appl
41	89	6.0	489	1	PCT-US02-36123-2614	Sequence 2614, Ap
42	89	6.0	560	1	PCT-US03-06415-2	Sequence 2, Appli
43	89	6.0	892	6	US-10-282-122A-70481	Sequence 70481, A
44	89	6.0	930	6	US-10-092-411A-5314	Sequence 5314, Ap
45	88.5	5.9	480	6	US-10-424-599-262034	Sequence 262034,

ALIGNMENTS

RESULT 1
US-10-424-599-227379
; Sequence 227379, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 227379
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_47352C.1.pep
US-10-424-599-227379

Query Match 44.7%; Score 666.5; DB 6; Length 268;
Best Local Similarity 51.1%; Pred. No. 9.4e-51;
Matches 134; Conservative 46; Mismatches 61; Indels 21; Gaps 5;

```
Qy         29 SRSSSSSSSAAPTVQATTSVHGHEEDPNQIPNNIRRQLPRSI-----TSSTSYKRFPLSR 83
              | :| |::   |        : |   |       ||:||:|:      :::|:   |
Db          6 SEASLSNNKKRP-------FNDHTHTP-------RRRLPKSLIALQHPNASSFSPHP-RP 50

Qy         84 CRAR-NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRK 142
              | :|   | |:| |:| ||:|   |:| | :|:::|  |  :     :| ||||:|:|||
Db         51 CDSRMTLPVMKFSGQISYSRTFDAVEKAATKLLQILQEKTTDMMQTAIGFDIEWKPTFRK 110

Qy        143 GVLPGKVATVQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDY 202
              || ||||| :||| |: :| |:|: ||||||:|| |:|| |::||| |||||:||:| ||
Db        111 GVPPGKVAVMQICGDTRHCHVLHLIHSGIPQNLQLLLEDPTVLKVGAGIDGDAVKVFRDY 170

Qy        203 GVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQ 262
                :|:| | |||  ||||:||| |||||||||| |: |:| |||:|||||||   |||:||:
Db        171 NISVKGVTDLSFHANQKLGGDHKWGLASLTEKLLSKQLKKPNKIRLGNWEAPVLSKEQLE 230

Qy        263 YAATDAYASWHLYKVLKDLPDA 284
              ||||||:||| ||: :||||||
Db        231 YAATDAFASWCLYQAIKDLPDA 252
```

RESULT 2
US-09-724-676-89006

```
; Sequence 89006, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 89006
;   LENGTH: 994
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-724-676-89006

  Query Match             23.6%;  Score 352;  DB 5;  Length 994;
  Best Local Similarity   37.7%;  Pred. No. 3.5e-22;
  Matches   75;  Conservative   41;  Mismatches   75;  Indels    8;  Gaps    3;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              : | : | | |:||  |::    :       |     |    || |:|| | : :| | |
Db         43 DLPFLEFTGSIVYSYDASDCSFLSE------DISMSLSDGDVVGFDMEWPPLYNRGKL-G 95

Qy        148 KVATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSI 206
              ||| :|:||  : | : |:    : || |: |:|:  : | |:||:||  ||  |: : :
Db         96 KVALIQLCVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKL 155

Qy        207 KDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
              |:  :|:|:||:|:   : | | || : |: |:|||   ||  ||  :||:: |  ||||
Db        156 KNFVELTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLYAAT 215

Qy        267 DAYASWHLYKVLKDLPDAV 285
              |||| : :|: |: | | |
Db        216 DAYAGFIIYRNLEILDDTV 234


RESULT 3
US-09-724-676A-89006
; Sequence 89006, Application US/09724676A
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676A
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 89006
;   LENGTH: 994
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-724-676A-89006

  Query Match             23.6%;  Score 352;  DB 5;  Length 994;
  Best Local Similarity   37.7%;  Pred. No. 3.5e-22;
  Matches   75;  Conservative   41;  Mismatches   75;  Indels    8;  Gaps    3;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              : | : | | |:||  |::    :       |     |    || |:|| | : :| | |
Db         43 DLPFLEFTGSIVYSYDASDCSFLSE------DISMSLSDGDVVGFDMEWPPLYNRGKL-G 95

Qy        148 KVATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSI 206
              ||| :|:||  : | : |:    : || |: |:|:  : | |:||:||  ||  |: : :
Db         96 KVALIQLCVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKL 155

Qy        207 KDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
              |:  :|:|:||:|:   : | | || : |: |:|||   ||  ||  :||:: |  ||||
Db        156 KNFVELTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLYAAT 215

Qy        267 DAYASWHLYKVLKDLPDAV 285
              |||| : :|: |: | | |
Db        216 DAYAGFIIYRNLEILDDTV 234


RESULT 4
US-09-949-001-22
; Sequence 22, Application US/09949001
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL000789
;  CURRENT APPLICATION NUMBER: US/09/949,001
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/231,323
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22
;   LENGTH: 1409
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-001-22

  Query Match             23.6%;  Score 352;  DB 5;  Length 1409;
  Best Local Similarity   37.7%;  Pred. No. 5.5e-22;
  Matches   75;  Conservative   41;  Mismatches   75;  Indels    8;  Gaps    3;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              : | : | | |:||  |::    :       |     |    || |:|| | : :| | |
Db         47 DLPFLEFTGSIVYSYDASDCSFLSE------DISMSLSDGDVVGFDMEWPPLYNRGKL-G 99

Qy        148 KVATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSI 206
              ||| :|:||  : | : |:    : || |: |:|:  : | |:||:||  ||  |: : :
Db        100 KVALIQLCVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKL 159

Qy        207 KDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
              |:  :|:|:||:|:   : | | || : |: |:|||   ||  ||  :||:: |  ||||
Db        160 KNFVELTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLYAAT 219

Qy        267 DAYASWHLYKVLKDLPDAV 285
              |||| : :|: |: | | |
Db        220 DAYAGFIIYRNLEILDDTV 238


RESULT 5
US-09-949-001-16
; Sequence 16, Application US/09949001
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL000789
;  CURRENT APPLICATION NUMBER: US/09/949,001
;  CURRENT FILING DATE:  2003-01-15
;  PRIOR APPLICATION NUMBER: 60/231,323
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 848
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 1432
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-001-16

  Query Match             23.6%;  Score 352;  DB 5;  Length 1432;
  Best Local Similarity   37.7%;  Pred. No. 5.7e-22;
  Matches   75;  Conservative   41;  Mismatches   75;  Indels    8;  Gaps    3;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              : | : | | |:||  |::    :       |     |    || |:|| | : :| | |
Db         43 DLPFLEFTGSIVYSYDASDCSFLSE------DISMSLSDGDVVGFDMEWPPLYNRGKL-G 95

Qy        148 KVATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSI 206
              ||| :|:||  : | : |:    : || |: |:|:  : | |:||:||  ||  |: : :
Db         96 KVALIQLCVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKL 155
```

```
Qy        207 KDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
              |:  :|:|:||:|:   : | | || : |: |:|||   ||  ||  :||:: |  ||||
Db        156 KNFVELTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLYAAT 215

Qy        267 DAYASWHLYKVLKDLPDAV 285
              |||| : :|: |: | | |
Db        216 DAYAGFIIYRNLEILDDTV 234


RESULT 6
US-10-374-077-71
; Sequence 71, Application US/10374077
;   GENERAL INFORMATION:
;        APPLICANT: Fu, Ying-Hui
;                   Yu, Chang-En
;                   Oshima, Junko
;                   Mulligan, John T.
;                   Schellenberg, Gerald D.
;        TITLE OF INVENTION: ANTIBODIES AGAINST GENE PRODUCTS RELATED TO
;                            WERNER'S SYNDROME
;        NUMBER OF SEQUENCES: 209
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Seed Intellectual Property Law Group
;             STREET: 701 Fifth Avenue, Suite 6300
;             CITY: Seattle
;             STATE: Washington
;             COUNTRY: USA
;             ZIP: 98104-7092
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/374,077
;             FILING DATE: 25-Feb-2003
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Rosenman, Stephen
;             REGISTRATION NUMBER: 43,058
;             REFERENCE/DOCKET NUMBER: 100107.401D1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (206) 622-4900
;             TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 71:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1432 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 71:
US-10-374-077-71

  Query Match             23.6%;  Score 352;  DB 6;  Length 1432;
  Best Local Similarity   37.7%;  Pred. No. 5.7e-22;
  Matches   75;  Conservative   41;  Mismatches   75;  Indels    8;  Gaps    3;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              : | : | | |:||  |::    :       |     || |:|| | : :| | |
Db         43 DLPFLEFTGSIVYSYDASDCSFLSE------DISMSLSDGDVVGFDMEWPPLYNRGKL-G 95

Qy        148 KVATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSI 206
              ||| :|:||  : | : |:    : || |: |:|:  : | |:||:||  ||  |: : :
Db         96 KVALIQLCVSESKCYLFHVSSMSVFPQGLKMLLENKAVKKAGVGIEGDQWKLLRDFDIKL 155

Qy        207 KDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAAT 266
              |:  :|:|:||:|:   : | | || : |: |:|||   ||  ||  :||:: |  ||||
Db        156 KNFVELTDVANKKLKCTETWSLNSLVKHLLGKQLLKDKSIRCSNWSKFPLTEDQKLYAAT 215

Qy        267 DAYASWHLYKVLKDLPDAV 285
              |||| : :|: |: | | |
Db        216 DAYAGFIIYRNLEILDDTV 234


RESULT 7
US-10-374-077-206
; Sequence 206, Application US/10374077
;   GENERAL INFORMATION:
;        APPLICANT: Fu, Ying-Hui
;                   Yu, Chang-En
;                   Oshima, Junko
;                   Mulligan, John T.
;                   Schellenberg, Gerald D.
;        TITLE OF INVENTION: ANTIBODIES AGAINST GENE PRODUCTS RELATED TO
;                            WERNER'S SYNDROME
;        NUMBER OF SEQUENCES: 209
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Seed Intellectual Property Law Group
;             STREET: 701 Fifth Avenue, Suite 6300
;             CITY: Seattle
;             STATE: Washington
;             COUNTRY: USA
;             ZIP: 98104-7092
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/374,077
;             FILING DATE: 25-Feb-2003
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Rosenman, Stephen
;             REGISTRATION NUMBER: 43,058
;             REFERENCE/DOCKET NUMBER: 100107.401D1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (206) 622-4900
;             TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 206:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1401 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 206:
US-10-374-077-206

  Query Match             22.1%;  Score 330;  DB 6;  Length 1401;
  Best Local Similarity   36.0%;  Pred. No. 4.9e-20;
  Matches   72;  Conservative   43;  Mismatches   71;  Indels   14;  Gaps    5;

Qy         88 NFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPG 147
              | | : | | |:||  |::       | : :  :  :  :  || |:|| | ::    ||
Db         37 NLPFLEFPGSIVYSYEASD----CSFLSEDISMRLSDGDV--VGFDMEWPPIYK----PG 86

Qy        148 K---VATVQICVDSNYCDVMHIFHSGI-PQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG 203
              |   || :|:||  : | : ||    : || |: |:|: :: | |:||:||  ||  |:
Db         87 KRSRVAVIQLCVSESKCYLFHISSMSVFPQGLKMLLENKSIKKAGVGIEGDQWKLLRDFD 146

Qy        204 VSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQY 263
              | ::   :|:|:||:|:   : | |  | : :: |:|||   ||  ||  :||:: |  |
Db        147 VKLESFVELTDVANEKLKCAETWSLNGLVKHVLGKQLLKDKSIRCSNWSNFPLTEDQKLY 206

Qy        264 AATDAYASWHLYKVLKDLPD 283
              |||||||   :|: | :| |
Db        207 AATDAYAGLIIYQKLGNLGD 226


RESULT 8
US-10-424-599-238476
; Sequence 238476, Application US/10424599
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
```

```
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 238476
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_57369C.1.pep
US-10-424-599-238476

  Query Match             17.0%;  Score 253.5;  DB 6;  Length 238;
  Best Local Similarity   39.9%;  Pred. No. 2.8e-14;
  Matches   59;  Conservative   28;  Mismatches   56;  Indels    5;  Gaps    4;

Qy        130 VGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHS-GIPQSLQHLIEDSTLVKVG 188
              |||||||||: ::   :   |||:|:|| :  | |  | ||  || ||   : |  :  ||
Db         85 VGLDIEWRPNTQRN-MQNPVATLQLCV-AERCLVFQILHSPSIPPSLVSFLADPNITFVG 142

Qy        189 IGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGG--DKKWGLASLTETLVCKELLKPNRI 246
              :||  |  ||  || ::: :| ||   | :::|    |: || ||   ::  |: || |:
Db        143 VGIQEDVEKLLEDYNLNVANVRDLRSFAAERLGDLELKRAGLKSLGLRVLGLEVAKPKRV 202

Qy        247 RLGNWEFYPLSKQQLQYAATDAYASWHL 274
                  |:   |: ||:|||| ||: |: :
Db        203 TRSRWDNPWLTAQQVQYAAVDAFLSYEI 230


RESULT 9
US-10-425-114-37706
; Sequence 37706, Application US/10425114
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 37706
;   LENGTH: 210
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 700346357_FLI.pep
US-10-425-114-37706

  Query Match             16.7%;  Score 249.5;  DB 6;  Length 210;
  Best Local Similarity   28.7%;  Pred. No. 5.5e-14;
  Matches   58;  Conservative   45;  Mismatches   90;  Indels    9;  Gaps    5;

Qy         80 PLSRCRARNFPAMRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPS 139
              |::  :  |   :|| | :: :           | ::|   |       ||||:|||||
Db         10 PMAATKVCN---VRFEGNVITTTVTASGAAVESWLDEILSVHRRRLHKLVVGLDVEWRPS 66

Qy        140 FRKGVLPGKVATVQICVDSNYCDVMHIFHSG-IPQSLQHLIEDSTLVKVGIGIDGDSVKL 198
              | :     | | ||:||    | :  : |:  :| :|   : |     ||:|:  |  :|
Db         67 FSRAY--SKTAIVQLCV-GRRCLIFQLLHADYVPNTLDEFLSDPDYTFVGVGVAADVERL 123

Qy        199 FHDYGVSIKDVEDLSDLANQKIGGD--KKWGLASLTETLVCKELLKPNRIRLGNWEFYPL 256
               :|| : : : |||::|| :::|    :  ||  :   ::  : ||  :| | |:   |
Db        124 ENDYDLEVANAEDLAELAAKEMGRPDLRNAGLQGIARAVMDAHVEKPQWVRTGPWDASSL 183

Qy        257 SKQQLQYAATDAYASWHLYKVL 278
              | :|::||  ||: |: : ::|
Db        184 SDEQIEYATIDAFVSFEVGRML 205


RESULT 10
US-10-425-114-71107
; Sequence 71107, Application US/10425114
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71107
;   LENGTH: 236
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB4762-022-E12_FLI.pep
US-10-425-114-71107

  Query Match             16.4%;  Score 245;  DB 6;  Length 236;
  Best Local Similarity   28.9%;  Pred. No. 1.6e-13;
  Matches   55;  Conservative   43;  Mismatches   86;  Indels    6;  Gaps    4;

Qy         92 MRFGGRILYSKTATEVDKRAMQLIKVLDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVAT 151
              :|| | :: :          | ::|   |       ||||:|||||| :     | |
Db         45 VRFEGNVITTTVTASGAAVESWLDEILSVHRRRLHKLVVGLDVEWRPSFSRAY--SKTAI 102

Qy        152 VQICVDSNYCDVMHIFHSG-IPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVE 210
              :|:||    | :  : |:  :| :|   : |     ||:|:  |  :| :|| : : : |
Db        103 LQLCV-GRRCLIFQLLHADYVPNTLDEFLSDPDYTFVGVGVAADVERLENDYDLEVANAE 161

Qy        211 DLSDLANQKIGGD--KKWGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDA 268
              ||::|| :::|    :  ||  :   ::  : ||  :| | |:   || :|::||  ||
Db        162 DLAELAAKEMGHPDLRNAGLQGIARVVMDAHVEKPQWVRTGPWDASSLSDEQIEYATIDA 221

Qy        269 YASWHLYKVL 278
              : |: : ::|
Db        222 FVSFEVGRML 231


RESULT 11
US-10-424-599-191095
; Sequence 191095, Application US/10424599
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 191095
;   LENGTH: 208
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_14578C.1.pep
US-10-424-599-191095

  Query Match             14.2%;  Score 212;  DB 6;  Length 208;
```

```
  Best Local Similarity   31.7%;  Pred. No. 1.1e-10;
  Matches   58;  Conservative   31;  Mismatches   60;  Indels   34;  Gaps    6;

Qy        120 TKRDESGIA----------------FVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDV 163
              |  |:|||                 |||| ||  : :  :   ||| :|:|:: | | :
Db         32 TITDKSGIVDKWIQVVSSTYAGKQRIVGLDTEWTTAKKPKM---KVAILQLCIE-NKCLI 87

Qy        164 MHIFH-SGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGG 222
              : :||   |||||:  : ||    ||:|:  |   | :|||:      |:| || :
Db         88 IQLFHMDNIPQSLRSFLMDSNFEFVGVGVINDLRMLKNDYGLECNKGIDVSLLAKE---- 143

Qy        223 DKKW-------GLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLY 275
              ||          |  | : ||  |: |   :    |:   |:: |::||  |||||: :
Db        144 --KWPHRISSGALKYLAKELVGLEMEKSKAVCTSEWQSKELTQTQIEYACIDAYASFKIG 201

Qy        276 KVL 278
              |::
Db        202 KMI 204


RESULT 12
US-10-424-599-229443
; Sequence 229443, Application US/10424599
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 229443
;   LENGTH: 205
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_49211C.1.pep
US-10-424-599-229443

  Query Match            12.6%;  Score 188;  DB 6;  Length 205;
  Best Local Similarity   60.4%;  Pred. No. 1.5e-08;
  Matches   32;  Conservative   12;  Mismatches    9;  Indels    0;  Gaps    0;

Qy        152 VQICVDSNYCDVMHIFHSGIPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGV 204
              :||| |: :| |:|: ||||||:|| |:|| |::||| |||||:||:| || :
Db          1 MQICGDTRHCHVLHLIHSGIPQNLQLLLEDPTVLKVGAGIDGDAVKVFRDYNI 53


RESULT 13
US-10-431-652-6289
; Sequence 6289, Application US/10431652
; GENERAL INFORMATION:
;  APPLICANT: Breton, Gary L.
;  APPLICANT:  Bush, David
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION:  BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH03-08
;  CURRENT APPLICATION NUMBER: US/10/431,652
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: US 09/328,352
;  PRIOR FILING DATE: 1999-06-04
;  PRIOR APPLICATION NUMBER: US 60/088,701
;  PRIOR FILING DATE: 1998-06-09
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6289
;   LENGTH: 223
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-10-431-652-6289

  Query Match            11.4%;  Score 169.5;  DB 6;  Length 223;
  Best Local Similarity   27.6%;  Pred. No. 7.1e-07;
  Matches   45;  Conservative   35;  Mismatches   72;  Indels   11;  Gaps    5;

Qy        129 FVGLDIEWRPSFRKGVLPGKVATVQICVDSNYCDVMHIFHSGIP--QSLQHLIEDSTLVK 186
              |:| | | :|:|: | :      :|:  :       ::||  :   : || :: :   :|
Db         58 FLGFDSESKPTFQVGEVSTGPHLIQLATEHK----AYLFHVNLSTLKFLQPILSNPKQIK 113

Qy        187 VGIGIDGDSVKLFHDYGVSIKDVEDLSDLANQKIGGDKKWGLASLTETLVCKELLKPNRI 246
              || |:  |   :||  |: ::   ||:       |  :: |:      |: : | |  ::
Db        114 VGFGLKNDK-HIFHKKGIELESCVDLAK-GFSHFGFKQQMGVQKAVALLLGQYLAKSKKV 171

Qy        247 RLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKD---LPDAVS 286
                ||   ||: ||: ||| ||||:  ::  ||    ||  :|
Db        172 GTSNWARKPLTSQQISYAAADAYAALLVFLELKKRKVLPIHIS 214


RESULT 14
US-10-425-114-71315
; Sequence 71315, Application US/10425114
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71315
;   LENGTH: 198
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: UC-ZMFLB73153A09_FLI.pep
US-10-425-114-71315

  Query Match            11.0%;  Score 163.5;  DB 6;  Length 198;
  Best Local Similarity   28.5%;  Pred. No. 2.1e-06;
  Matches   47;  Conservative   33;  Mismatches   74;  Indels   11;  Gaps    5;

Qy        118 LDTKRDESGIAFVGLDIEWRPSFRKGVLPGKVATVQICVD--SNYCDVMHIFH---SGIP 172
              |:   |      :|:| ||||:| |     ||: :||  |  :   |:: ::      :
Db          1 LNATSDIEACKIIGMDCEWRPNFEKNTKSSKVSIIQIASDKIAFIFDLIKLYEDDPKALD 60

Qy        173 QSLQHLIEDSTLVKVGIGIDGDSVKLFHDYG--VSIKDVEDLSDLANQKIGGDKKWGLAS 230
                |: ::  | ::|:|  |  |  :|   ||     :  | : |:  ||:       ||:
Db         61 SCLRRVMCSSKILKLGYDIQCDLHQLTRSYGELECFQSYEMVLDM--QKLFKGVTGGLSG 118

Qy        231 LTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLY 275
              |:: ::   | |    |  |||  ||:: | :||| ||    |::
Db        119 LSKEILGAGLNKTR--RNSNWEQRPLTQNQKEYAALDAVVLVHIF 161


RESULT 15
US-10-276-774-1634
; Sequence 1634, Application US/10276774
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  APPLICANT:  Tang, Y, Tom et al
;  TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
;  FILE REFERENCE: 21272-030
;  CURRENT APPLICATION NUMBER: US/10/276,774
;  CURRENT FILING DATE:  2002-11-18
;  PRIOR APPLICATION NUMBER: 09/560,875
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 09/496,914
;  PRIOR FILING DATE: 2000-02-03
;  NUMBER OF SEQ ID NOS: 2700
```

```
;  SOFTWARE: Custom
; SEQ ID NO 1634
;   LENGTH: 127
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-276-774-1634

  Query Match             10.1%;  Score 151;  DB 6;  Length 127;
  Best Local Similarity   35.0%;  Pred. No. 1.5e-05;
  Matches   41;  Conservative   20;  Mismatches   48;  Indels    8;  Gaps    2;

Qy        171 IPQSLQHLIEDSTLVKVGIGIDGDSVKLFHDYGVSIKDVEDLSDLA-----NQKIGGDKK 225
              :|::|  :: | |::|||:|   |: ||  |||: ::   ||  ||     |    |
Db          3 LPRTLLDILADGTILKVGVGCSEDASKLLQDYGLVVRGCLDLRYLAMRQRNNLLCNG--- 59

Qy        226 WGLASLTETLVCKELLKPNRIRLGNWEFYPLSKQQLQYAATDAYASWHLYKVLKDLP 282
               | || ||::   | |   :|  ||:   |:: |: ||| ||  |  |:  |   |
Db         60 LSLKSLAETVLNFPLDKSLLLRCSNWDAETLTEDQVIYAARDAQISVALFLHLLGYP 116


Search completed: June  6, 2003, 10:31:03
Job time : 43 secs
```